TOPICS AND GEOGRAPHICAL DIFFUSION OF KNOWLEDGE IN TOP ECONOMIC JOURNALS

Published version:

DOI:10.1111/ecin.12815

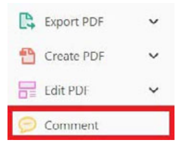
WILEY

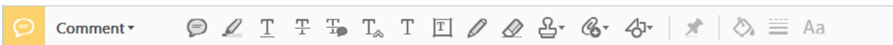
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

Required software to e-Annotate PDFs: Adobe Acrobat Professional or Adobe Reader (version 11 or above). (Note that this document uses screenshots from Adobe Reader DC.)
The latest version of Acrobat Reader can be downloaded for free at: http://get.adobe.com/reader/

Once you have Acrobat Reader open on your computer, click on the Comment tab (right-hand panel or under the Tools menu).

This will open up a ribbon panel at the top of the document. Using a tool will place a comment in the right-hand panel. The tools you will use for annotating your proof are shown below:



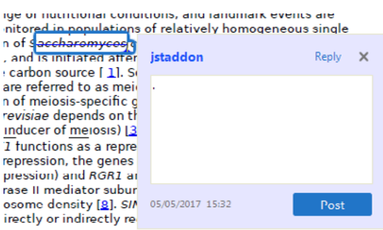


1. Replace (Ins) Tool – for replacing text.

Strikes a line through text and opens up a text box where replacement text can be entered.

How to use it:

- Highlight a word or sentence.
- Click on .
- Type the replacement text into the blue box that appears.

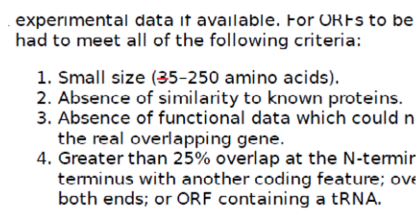


2. Strikethrough (Del) Tool – for deleting text.

Strikes a red line through text that is to be deleted.

How to use it:

- Highlight a word or sentence.
- Click on ..
- The text will be struck out in red.

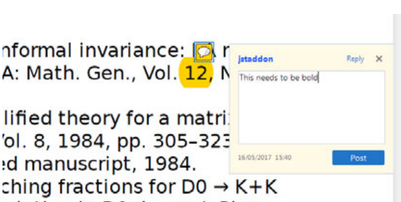


3. Commenting Tool – for highlighting a section to be changed to bold or italic or for general comments.

Use these 2 tools to highlight the text where a comment is then made.

How to use it:

- Click on .
- Click and drag over the text you need to highlight for the comment you will add.
- Click on .
- Click close to the text you just highlighted.
- Type any instructions regarding the text to be altered into the box that appears.

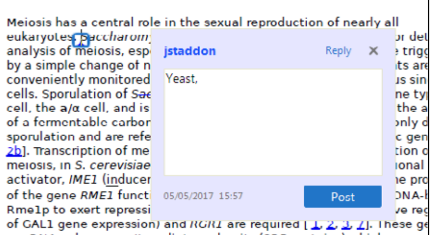


4. Insert Tool – for inserting missing text at specific points in the text.

Marks an insertion point in the text and opens up a text box where comments can be entered.

How to use it:

- Click on .
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the box that appears.



WILEY

USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

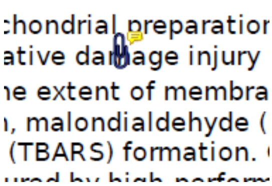
5. Attach File Tool – for inserting large amounts of text or replacement figures.

Inserts an icon linking to the attached file in the appropriate place in the text.

How to use it:

- Click on .
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.

The attachment appears in the right-hand panel.

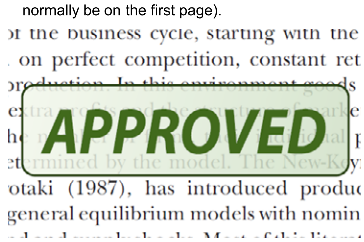


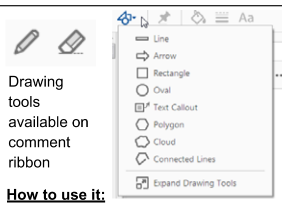
6. Add stamp Tool – for approving a proof if no corrections are required.

Inserts a selected stamp onto an appropriate place in the proof.

How to use it:

- Click on .
- Select the stamp you want to use. (The Approved stamp is usually available directly in the menu that appears. Others are shown under *Dynamic*, *Sign Here*, *Standard Business*).
- Fill in any details and then click on the proof where you'd like the stamp to appear. (Where a proof is to be approved as it is, this would normally be on the first page).





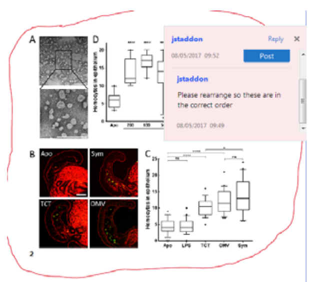
Drawing tools available on comment ribbon

7. Drawing Markups Tools – for drawing shapes, lines, and freeform annotations on proofs and commenting on these marks.

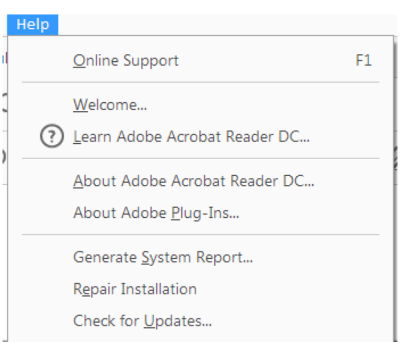
Allows shapes, lines, and freeform annotations to be drawn on proofs and for comments to be made on these marks.

How to use it:

- Click on one of the shapes in the Drawing Markups section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, right-click on shape and select *Open Pop-up Note*.
- Type any text in the red box that appears.



For further information on how to annotate proofs, click on the Help menu to reveal a list of further options:



AUTHOR QUERY FORM

Dear Author,
During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof.
Many thanks for your assistance.

Query References	Query	Remarks
Q1	Please confirm that given names (blue) and surnames/family names (vermilion) have been identified and spelled correctly	
Q2	Please check if link to ORCID is correct.	
Q3	Please provide fax numbers for all authors. Please provide job titles for authors Magda Fontana and Paolo Racca. Please check whether inserted affiliation details are appropriate for all affiliations.	
Q4	Please define IF.	
Q5	Please check whether the phrase “applied and applied theory papers“ conveys intended meaning.	
Q6	Please note that Gibson et al. (2017), Frey et al. (2009) has been cited in text but not provided in list. Please provide in list or delete the citation.	
Q7	Please note that as per journal style equations have to be linearized. We have not linearized to avoid introducing errors.	
Q8	Please note that endnote 9 is provided without text. Please supply appropriate text.	
Q9	Reference "Blei & Lafferty, 2009" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q10	Please provide journal title for ref. Blei and Lafferty (2009).	
Q11	Reference "Hoffman et al, 2010" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	
Q12	Reference "Řehůřek & Sojka, 2010" is not cited in the text. Please indicate where it should be cited; or delete from the reference list.	

Funding Info Query Form

Please confirm that the funding sponsor list below was correctly extracted from your article: that it includes all funders and that the text has been matched to the correct FundRef Registry organization names. If no FundRef Registry organization name has been identified, it may be that the funder was not found in the FundRef registry, or there are multiple funders matched in the FundRef registry. If a name was not found in the FundRef registry, it may not be the canonical name form, it may be a program name rather than an organization name, or it may be an organization not yet included in FundRef Registry. If you know of another name form or a parent organization name for a "not found" item on this list below, please share that information.

Funding Agency	**FundRef Organization Name**
Collegio Carlo Alberto	Collegio Carlo Alberto

TOPICS AND GEOGRAPHICAL DIFFUSION OF KNOWLEDGE IN TOP ECONOMIC JOURNALS

MAGDA FONTANA, FABIO MONTOBBIO and PAOLO RACCA*

We study the evolution of topics in economics and their geographical specialization by analyzing 13,233 papers from seven top journals between 1985 and 2012 and their forward citations. The share of U.S. publications declines from 75% to 64% with a corresponding increase of the European share from 12% to 24%. We use topic modeling and document the evolution of the discipline over 27 years. We estimate, with a quasi-structural model, the citation lag distribution for 18 different topics and three large geographical areas. The modal citation lag is about 6.7 years in the entire sample and 4.8 years for citations from the top 100 journals. We quantify (1) the home bias effect in citations, (2) how it fades away over time, (3) the long lasting impact of U.S. publications vis-à-vis other geographical areas and (4) the higher speed of diffusion and faster obsolescence in the United States. (*JEL* A14, I23, 033, A11)

I. INTRODUCTION

The creation and diffusion of scientific knowledge have a great impact on economic prosperity of countries and regions (Grossman and Helpman 1991; Phelps 1996; Romer 1991) and the geographic location of top scientific research and its rate of spatial diffusion has important implications for the evolution of science and for science policy. In the policy arena, public support of scientific research emphasizes the role of excellence in science. The economic benefits of this public support depend upon the fruits of this research, the ability to stay ahead in research, and to learn from excellence. Thus for both modeling science evolution and research policy purposes, it is important to understand the geographic and temporal dimensions of the spread of newly created scientific knowledge and the specific evolution of the different fields.

We tackle this issue studying scientific progress in economics. Exploiting the increased availability of large bibliometric databases, a set of recent papers has provided some quantitative evidence on the relative growth of different fields in economics and the degree of geographic

*The authors are grateful to the participants of the Fourth Geography of Innovation Conference, Barcelona, 2018 and to the 12th Workshop on the Organization, Economics and Policy of Scientific Research. They are also grateful to Bath, 2018, for valuable comments on a previous version of the paper. The authors are also indebted to Gianluca Tarasconi and Martina Iori for their invaluable research assistance and with Andrea Bonaccorsi, Paul David, Marco Guerzoni, Paula Stephan and Reinhilde Veugelers for illuminating discussions on the issues covered in the paper. They are grateful to Marco Guerzoni for helping us with the access to the JSTOR database. Finally, authors thank the anonymous reviewers for their careful reading of the manuscript and for their insightful suggestions. The usual disclaimer applies.

Fontana: xxx, Department of Economic and Statistics "Cognetti De Martiis", University of Turin, Torino 10153, Italy. DESPINA Big Data Lab, University of Turin, Torino 10123, Italy.Phone +39 011 670 3888, E-mail magda.fontana@unito.it

Montobbio: Associate Professor, Department of Economic Policy, Università Cattolica del Sacro Cuore, Milan 20123, Italy. iCRIOS, Bocconi University, Milan 20136, Italy. BRICKS, Collegio Carlo Alberto, Turin 10024, Italy. Phone +39 02 72342921, E-mail fabio.montobbio@unicatt.it

Racca: xxx, DESPINA Big Data Lab, University of Turin, Torino 10123, Italy. Phone +39 011 6703101, E-mail paolo.racca@unito.it

ABBREVIATIONS

AER: American Economic Review
EU: European Union
IER: International Economic Review
IF: XXX
JPE: Journal of Political Economy
LDA: Latent Dirichlet Allocation
QJE: Quarterly Journal of Economics
RME: Relatively More Empirical
RMT: Relatively More Theoretical
RES: Review of Economic Studies
RESTAT: Review of Economic and Statistics
RoW: Rest of the World
RSA: Relative Scientific Advantage

Economic Inquiry
(ISSN 0095-2583)

doi:10.1111/ecin.12815

concentration of publications in top journals (Angrist et al. 2017; Card and DellaVigna 2013; Claveau and Gingras 2016; Hamermesh 2013, 2018; Kim, Morse, and Zingales 2006; Kosnik 2015). The general results are the growing importance of empirical vis-à-vis theoretical work concerning most of the different fields within economics. In addition even if scientific knowledge is typically treated as codified knowledge that diffuses quickly in the global network of scientists, excellence in economics remains highly concentrated and there is scant evidence on the rate of geographical diffusion of different fields in economics (Anauati, Galiani, and Gálvez 2016, 2018; Galiani and Gálvez 2017; Hargreaves Heap and Parikh 2005).

This paper contributes to the growing body of literature that quantitatively analyzes the rate of diffusion and obsolescence of different fields in the economic discipline looking at the papers' characteristics and their citation performance (Anauati, Galiani, and Gálvez 2016, 2018; Angrist et al. 2017; Galiani and Gálvez 2017). We estimate precisely, using a quasi-structural model, the life cycle of the papers taking into account their topic, and the geographical origin and cohort of both citing and cited papers.

First, we ask which topics economists are researching, and which ones are represented in a set of top journals, using topic modeling that provides some advantages with respect to the more commonly used JEL codes. Secondly, we ask how these topics are distributed across geographical areas, studying whether there is specialization in producing knowledge related to a given topic. Finally, we estimate the speed of diffusion and decay of knowledge in economics within and among all combinations of geographical areas and we explore which topics diffuse more rapidly and are more influential. In particular, we estimate the citation lag distributions and describe the citation patterns among all combinations of three large geographical areas United States, Europe, and Rest of the World (RoW).

This paper starts from the analysis of 13,233 focal papers from seven top journals in economics (Conroy and Dusansky 1995). We study the papers by topic and geographical area, eliciting the thematic structure of the articles through topic modeling analysis on full-texts (Latent Dirichlet Allocation, LDA; Blei, Ng, and Jordan 2003). Papers and topics are then assigned to countries and geographical areas via the authors' affiliations. The paper exploits two sets of citations to these focal papers. The first one includes all 780,180 citations from 1985 to 2015. The second one is a restricted sample of 227,000 citations coming from the top 100 journals in the field (Guerrero-Bote and Moya-Anegon 2012). We analyze the process of diffusion and obsolescence of knowledge contained in the papers estimating the citation lag distribution for 18 different topics and three large geographical areas. To perform this task we adopt a quasi-structural model as proposed by Caballero and Jaffe (1993) and discussed in Jaffe and Trajtenberg (1996) and Hall, Jaffe, and Trajtenberg (2001) for patent data. It combines two exponentials to model the likelihood of citations taking into account different attributes of the cited and citing publications.

Our results can be summarized as follows. There is a prevalence of papers from researchers affiliated in the United States. This prevalence declines between 1985 and 2012 from 75% to 64% with a corresponding increase of the European share, which approaches one fourth of the papers at the end of the observation period. The estimated shape of the citation lag distribution in economics shows that the modal lag on average is about 6.7 years in the entire sample and 4.8 years in the restricted sample. Citations to articles in top journals in economics have a slow rate of decay. On average after 30 years the estimated probability to be cited is still 46% of its maximum value.

Our estimations quantify precisely four overlapping effects: (1) there is a home bias effect in citations. For example, a publication originated in Europe is 39% more likely to get a citation from an average European publication than is a random U.S. publication. (2) This effect fades away over time. We find that the probability that a publication in Europe or RoW would cite—1 year after the publication date—a publication originated in the United States is 40% and 33%, respectively, lower than citations originated in the United States, but 30 years later the figures turn out to be 21% and 16% higher. (3) There is a long lasting impact of U.S. publications vis-à-vis other geographical areas. Papers from Europe and the RoW relatively cite more U.S. papers and these citations come with a longer lag. (4) There is a higher speed of diffusion and faster obsolescence in the United States. Citations in the United States come faster and show a higher rate of decay. These results do not depend upon the ranking of the citing journals and give a precise quantitative expression to commonly held perceptions about the dynamism of the economic discipline in the United States vis-à-vis other countries (Cardoso, Guimaraes, and Zimmermann 2010; Frey and Eichenberger 1993).

Finally, we find that there are different diffusion and decay path for different topics: some topics (like Growth and Technology) are highly cited during the first years but have a quick obsolescence, and other topics like Business Finance and Banks and Education display relatively lower obsolescence rates. We show, however, that the diffusion and decay rate of the different topics are different if we consider also the citing papers in the top 100 journals. This has important consequences for citation-based indicators; the differences across fields in impact factors, calculated on the first years after publication (as emphasized by Anauati, Galiani, and Gálvez 2016, 2018), are also affected by the type of citing journals considered.

Our paper is divided into six sections. Section II briefly surveys the available evidence and discusses the novelty of the paper. Section III explains the model. Section IV describes the data and the methodology. Section V shows the patterns of geographical specialization and topic evolution. Section VI gives the econometric results and provides a discussion of the limitations and of the interpretative framework. Section VII concludes.

II. BACKGROUND AND MOTIVATION

Recent papers study the evolution of the different fields in economics using different samples and methodologies. Angrist et al. (2017) analyze 134,892 papers published in 80 journals between 1980 and 2015. They build their field classification on JEL codes, titles, and keywords, the publishing journal and, finally, the fields of the papers that a paper cites and use machine-learning and clustering algorithms on a trained dataset. They also use a machine learning algorithm to assign three styles to papers: theoretical, empirical, or econometrics. Hamermesh (2013) skimmed 748 articles published in the American Economic Review (AER), Journal of Political Economy (JPE), and Quarterly Journal of Economics (QJE), in 1963, 1973, 1983, 1993, 2003, and 2011 classifying the papers according to five research methodologies: theory, theory with simulation, empirical using borrowed data, empirical using self-generated data, and experiment. Kim, Morse, and Zingales (2006) mainly use JEL codes on a set of 146 articles with over 500 cites from 41 top economic journals. Card and DellaVigna (2013) use JEL codes in the articles of the top five journals.

Angrist et al. (2017) show that the publication shares for labor and industrial organization have declined since the mid-late 1980s. Also a miscellaneous category is showing a greater impact in recent years. It includes various fields like environmental economics, experimental economics, urban economics, and political economy. Kim, Morse, and Zingales (2006) find an increasing importance among the highly cited articles of growth and development and a large weight for finance and econometrics. Card and DellaVigna (2013) looking at the 13,089 papers published in the top five journals from 1970 to 2012 find that the relative shares of the different fields are fairly constant over time. Kosnik (2015) analyzes 20,321 papers published in seven top-tier journals from 1960 to 2010 showing that, while most fields have retained a stable importance, pure macroeconomics has experienced a significant decrease in importance over time in favor of a growing interest in the microeconomic foundations of macroeconomics.

Claveau and Gingras (2016) adopt an unsupervised procedure that combines bibliometrics and networks analysis to study the dynamics the fields in economics from 1956 to 2014 on a sample of 450,000 papers drawn from Thomson Reuter's Web of Science. They detect the disappearance of the field dedicated to general economic theory in the late 1970 and, in the early 1990, the dissolution of the formerly cohesive field of econometrics in several specialties centered on specific methods.

Finally, Kelly and Bruestle (2011) do not focus only on the top journals and analyses 525,956 articles in 1,373 peer-reviewed journals from 1969 to 2007 from the ECONLIT database. They find significant changes in the percentage share of the different subjects in economics with an increase of specialty journals. In particular, in partial contradiction with Angrist et al. (2017), they find that Finance, Development, and Industrial Organization significantly increased their shares in the 40 years considered. At the same time Macroeconomics, Microeconomics and Labor declined.

Recent evidence also suggests that publications in the top journals come largely from the United States. Hamermesh (2013) shows that for his sample the share of United States-/Canada-based authors fell from 92% in 1963–1993 to 83% in 2003 and 2011. Kim, Morse, and

Zingales (2006) show that 85% of the most-cited papers originated in U.S. institutions. They find also that this share does not decline over time. In the top journals it is also more likely to publish a paper on the United States. Considering only the top five journals, Das et al. (2013) find a strong U.S. premium in this respect. This corroborates Bardhan's (2003) concerns about a possible misallocation of talent across research institutions and a diversion of research incentives away from the study of other countries.[1]

Scant evidence is, however, available on how scientific knowledge diffuses across space. Kim, Morse, and Zingales (2009) find that affiliation with a top 25 universities in the United States generates a positive marginal effect in term of research productivity in the 1970s and in the 1980s. This effect disappears in the 1990s. This decline is explained by the reduced importance of physical access to productive research colleagues, due to innovations in communication technology. However despite this reduced localization effect (i.e., university fixed effects), they find that elite universities have a higher average productivity because of agglomeration of top researchers with high research reputation.[2] Kalaitzidakis et al. (2004) still find positive spillovers from links to U.S. departments. They look at the activities of economics departments in Europe from 1993 to 1998 using survey data finding that faculties that have connections with North American departments (visiting programs, education received in North America by European faculty, and co-authorship) have higher research output and productivity (in terms of published pages) in 10 core journals.

Finally Anauati, Galiani, and Gálvez (2016) study the life cycle of economic papers across fields of in economics. They exploit 9,672 articles in the top five economic journals (1970–2000) and citations data obtained from Google Scholar. They show that papers display a life cycle: there is a diffusion path, a peak in terms of citations and then an obsolescence process. They analyze four fields (applied, applied theory, econometric methods, and theory) and find that applied and applied theory papers—relative to theoretical papers—receive more yearly citations in the first years following publication and have a longer lifespan. In addition Anauati, Galiani, and Gálvez (2018) analyze citations patterns across different journal tiers. They find that on average articles in nontop five journals receive less citations and have a faster obsolescence. So articles in the top five journals have a longer life cycle. However, they find that the differences in overall citations patterns across journal tiers change across fields and depend upon which articles' citation quantile is taken into consideration.

A. Knowledge Structure, Diffusion, and Citations

This paper aims at finding patterns in knowledge production and diffusion across geographical areas. First, it describes the main topics studied in economics. We use topic modeling on full texts to assign a set of topics to each paper. Second, it shows how different geographical areas are positioned in terms of these topics. The underlying idea is that countries might exhibit specialization in producing knowledge related to a given topic. Topics are assigned to countries and geographical areas via the authors' affiliations (e.g., the address of the institution where they are employed or to which they are affiliated).

Thirdly, it focuses on how the knowledge produced in a specific location circulates among geographical areas. The process of diffusion of scientific knowledge across geographical areas is accounted for by controlling for the effects of truncation, changes in citation patterns, and topic effects. In particular, we explore for the first time the citation patterns among all combinations of three large geographical areas United States, Europe, and RoW. This paper provides a picture of the geographic dimension of citation diffusion, by examining the extent and speed of diffusion of citations within and among all combinations of these geographical areas. We estimate the extent and nature of localization of citations within each of these geographical areas, analyze differences among the geographical areas in their absorption of external knowledge and, finally, map significant pairwise effects.

In order to do so, we exploit citations of previous work in scientific articles. The scientific community is regulated by a set of norms and rules

1. Relatedly when they look not only at the top journals but, more generally, at a large database that includes 76,046 empirical economics papers published between 1985 and 2005, Das et al. (2013) find that the number of research articles on a given country increase with the country's population and wealth. In fact they find a strong correlation between per-capita research output and per-capita GDP.

2. In general the higher scientific productivity in top universities depends upon the ability to attract and retain productive and motivated scientists. However, using university fixed effects Kim, Morse, and Zingales (2009) identify the average individual productivity at the top schools, due to a potentially positive marginal effect of the top universities on their faculty.

guiding the behavior of researchers (Dasgupta and David 1994; Stephan 2012). One important norm is to cite previous work to establish scientific credit and to identify scientific antecedents (Hamermesh 2018; Kuhn 1962; Merton 1968); citations, as shown in the previous section, measure the impact and quality of scientific findings and, by extension, of a researcher, an institution, or a journal. Citations also affects knowledge creation and diffusion more indirectly: most of the metrics used to evaluate researchers and research institutions and their grant applications are based on citation counts (e.g., Hamermesh 2018; Gibson, Anderson, and Tressler 2014, 2017; Ellison 2013; Hamermesh and Pfann 2012).

Also many studies on technological knowledge rely on citation data also to analyze the diffusion of scientific ideas, the creation and evolution of scientific networks, and the role of top scientists and inventions (e.g., Breschi and Lissoni 2009; Fleming, King, and Juda 2007; Gittelman and Kogut 2003; Hall, Jaffe, and Trajtenberg 2005; Jaffe and Trajtenberg 1999; Narin, Hamilton, and Olivastro 1997; Singh 2005; Trajtenberg 1990). Our assumption is that a scientific publication is a proxy for a new bit of knowledge and citations to previous work signal whether a specific bit of knowledge is used in the construction of a new bit. So we exploit the probability of citation as a proxy for the probability of useful knowledge flow, which we measure with empirical citation frequencies.

We analyze how the probability of citation is affected by the time, geographic location, and scientific topic of each paper and by the specific relationship between the characteristics of the citing and cited papers. We calculate the rate of diffusion and decay in different locations and for different topics and, in particular, we measure the localization of scientific citations and how these localization effects change over time. There is an enormous amount of empirical evidence on localization of technological knowledge (e.g., Bottazzi and Peri 2003; Breschi and Lissoni 2009; Criscuolo and Verspagen 2008; Jaffe and Trajtenberg 1999; Jaffe, Trajtenberg, and Henderson 1993; Maruseth and Verspagen 2002; Peri 2005). Our assumption is that, far from freely diffusing in space without obstacles, also scientific knowledge circulation shows localization patterns; in parallel, we expect that the localization effects could fade away over time. In this paper, we draw from the literature on patent citations and exploit information on both citing and cited papers. We estimate the probability (and the changes over time of this probability) that a particular group of scientists (the citing ones) in a specific location and year will benefit from some other group of scientists (the cited ones) active on a specific topic in a specific location and year.

We assume that this probability is determined jointly by the characteristics of each group, and the nature of the relationship between the groups. In addition, scientific papers become obsolete. The diffusion path is therefore dependent upon the time lag between the citing and the cited papers and the outcome is the combination of the diffusion and obsolescence process. We expect that the citation probability first rise and then fall with elapsed time and this paper also provides and attempt to estimate exactly the citation lag distribution (Anauati, Galiani, and Gálvez 2016; Galiani and Gálvez 2017). In doing so, it is also necessary to take into account that the propensity to publish and the propensity to cite vary over time and space.

III. THE MODEL

To explore variations across topics and geographical areas of the propensity to cite, we exploit a quasi-structural model as proposed by Caballero and Jaffe (1993) and discussed in Jaffe and Trajtenberg (1996, 1999), Hall, Jaffe, and Trajtenberg (2001), and Bacchiocchi and Montobbio (2010). A full discussion of its derivation can be found in Caballero and Jaffe (1993) in the context of the production of new technological ideas (patents). We apply it to analyze the field of economics where the new bit of knowledge produced is a scientific paper. Summing up the points raised in Section II we assume that a citation is observed when the author has read the paper. If he/she has not discovered a better article, he/she will cite the paper, establishing scientific credit and identifying prior useful work. Researchers take time in seeing others' papers. This generates a diffusion lag that is affected by geography and fields effects. On the other hand, over time, the probability of a paper being read and cited decreases because new articles that are published could replace it. So the probability of citation is proportional to the probability of the article being read and not supplanted and, as a result, depends upon its importance and on how far the field has moved on.

These factors can be captured by a citation function that has two main components: diffusion and obsolescence. In particular, we model the citation function *p(k, K)*—the likelihood for a publication K in year T to cite a publication k in year t—combining two exponentials:

$$(1)\ p(k,K) = \alpha(k,K)\exp\left[-\beta_1(k,K)(T-t)\right] \times \left[1-\exp\left[-\beta_2(T-t)\right]\right]$$

The second and the third factors in Equation ((1) determine, respectively, the processes of obsolescence and diffusion over time that depend upon the citation lag $(T-t)$ between the citing and the cited paper.[3] The rate of diffusion is determined by β_2 (greater β_2 means faster diffusion), while the obsolescence rate is determined by $\beta_1(k, K)$ (greater β_1 means faster obsolescence) (see also endnote 5). The dependence of this term on k and K means that it depends upon attributes of both the citing and the cited items. The same stands also for the multiplicative term $\alpha(k, K)$.

In order to capture the joint effect of these three terms on the shape of the function, it is convenient to refer to the modal lag, that is, the lag value which maximizes the function. It is equal to $(1/\beta_2) * \log(1+\beta_2/\beta_1)$ and quantifies after how much time the publication is more likely to be cited. Another useful measure is the integral from zero to infinity of Equation (1) with respect to the lag. This *cumulative probability*, equal to $(\alpha\beta_2)/[\beta_1(\beta_1+\beta_2)]$ (note that it is proportional to the multiplicative factor α), is an estimation of the expected number of citations that a single publication will receive from one random publication per year forever.

Following Caballero and Jaffe (1993) the underlying idea of Equation ((1) is that the citation equation can be seen as a component of a research productivity parameter that depends upon the stock of existing knowledge. Caballero and Jaffe (1993) apply this framework to measure research productivity in the context of an endogenous growth model with quality ladders. We extend this idea to the production of knowledge in a specific scientific discipline. Similarly to what has been done with patent data, we also extend the analysis to a multicountry, multifield context. This finer structure allows to analyze, for example, whether Europeans are slower to pick up knowledge produced in the United States or whether different fields display differences in the process of knowledge diffusion and decay. In particular in this paper we follow Bacchiocchi and Montobbio (2010) and use the following specification:

$$(2)\ p_{t,a,\text{topic},T,A} = \frac{c_{t,a,\text{topic},T,A}}{\left(n_{t,a,\text{topic}}\right)\left(n_{T,A}\right)}$$
$$= \alpha_{\text{const}}\alpha_t\alpha_{\text{topic}}\alpha_T\alpha_{aA} \times \exp\left[-\beta_{1\text{const}}\beta_{1\text{topic}}\beta_{1aA}(T-t)\right] \times \left[1-\exp\left[-\beta_2(T-t)\right]\right] + \epsilon_{t,a,\text{topic},T,A}$$

where t and T are publication years of the focal and citing papers, a and A are the macro-areas of the focal and citing papers and *topic* refers to the topic of the focal papers. Hence, $c_{t,\,a,\,\text{topic},\,T,\,A}$ is the amount of citations received by the papers on a specific *topic*, in a specific location a and in year t from papers published in year T originating in area A. Similarly, $n_{t,\,a,\,\text{topic}}$ is the amount of papers in the *(t, a,* topic*)*-group and $n_{T,\,A}$ the amount of papers in the *(T, A)*-group of citing papers.[4] Therefore, our $p_{t,\,a,\,\text{topic},\,T,\,A}$ can be interpreted as a proxy of the likelihood of a *(t, a,* topic*)*-paper to receive a citation from a *(T, A)*-paper. If the error term $\varepsilon_{t,\,a,\,\text{topic},\,T,\,A}$ is well-behaved, this model can be estimated by nonlinear least squares.

In this specification, the term $\alpha(k, K)$ has been factorized as product of a fixed coefficient, of effects of single categorical variables (t, topic and T) and of an interaction effect between geographical categorical variables (a and A). For the corresponding parameters to be identifiable, all these effects have a base case value of 1. Therefore, the interpretation of these parameters is relative to their own base case. If, for instance, *topic 0* is the base case for α_{topic} (so that $\alpha_{\text{topic}=0}$ is constrained to unity), and $\alpha_{\text{topic}=1} = 1.2$, this would imply, ceteris paribus, that *topic 1* is 20% more likely to be cited than *topic 0*. The same reasoning holds for the α_{aA} term too, but this time the base case corresponds to a pair of focal-forward areas. In fact, α_{aA} captures, in average terms, the relative likelihood that a paper from area a gets cited from a paper from area A. Analogous considerations hold for the factorization of the obsolescence term $\beta_1(k, K)$. For instance, a $\beta_{1,\,\text{topic}=i}$ significantly greater than 1 indicates a relatively faster obsolescence rate for topic i with respect to the base case.[5]

3. In what follows we also use the term focal papers to refer to the cited papers.

4. Please note that in our empirical work we did not have the analogous quantity for all potentially citing papers.

5. It can be noted that increases in β_2 (holding β_1 constant) tend to increase the overall citation intensity. For example the impact of increases of β_2 on the cumulative distribution is very similar to the impact of α. Indeed, faster diffusion, holding obsolescence constant, generates a change in

IV. DATA AND METHODOLOGY

The dataset combines data from two different sources: the ISI—Web of Science database, used for bibliographic information, citations and authors affiliations, and the JSTOR Digital Library, that contains the full text of articles (details of the record linkage procedure in Section A2 in the Appendix).

The starting point of this paper is a focal set of documents that includes the articles published in the so-called Blue Ribbon Eight journals[6] (Conroy and Dusansky 1995). Our analysis does not include the Journal of Economic Theory (JET), because the full text was not available in JSTOR. Due to the coverage constraints of the original data sources, the time period is limited to 1985–1996 for the JPE, and to 1985–2012 for the other six journals. In our study we consider articles, notes and proceedings papers.[7]

Table 1 shows the number of focal documents used in the analysis, grouped by journal. Our sample covers the 97% of the documents published in the periods specified above. From each document, we retrieve geographical areas from affiliations and topics from the full text. In particular, we use the addresses of author affiliations (e.g., the address of the institution where they are employed or to which they are affiliated) provided by the Web of Science to characterize documents in terms of geographical area (United States, Europe, and RoW). Articles with multiple affiliations are attributed to each area with the appropriate fraction (details in Section A3 in the Appendix).

TABLE 1
Number of Documents

Journal	1985–1999	2000–2012
AER	2,484	2,445
ECON	896	747
IER	749	634
JPE	677	—
QJE	710	525
RES	613	570
RESTAT	1,297	886
Total	7,426	5,807

We adopt LDA (Blei, Ng, and Jordan 2003), a standard topic modeling tool, to extract the thematic structure from the full text of the articles. This means that, through an unsupervised procedure, we characterize each article in terms of its most representative themes (see Section A1 in the Appendix for more details). In the LDA a topic is defined as a probability distribution over a vocabulary; in particular, one assumes that documents have been generated from x topics and that every document can contain more than one topic in different proportions. Specifically, topics are distributions over the words of the vocabulary, drawn from a uniform Dirichlet distribution.

Topic modeling provides a mapping that is more stable and reliable than grouping according to JEL code since it is not affected by changes in classification (for a history of JEL codes see Cherrier 2017) and is not biased by author strategic self-attribution of codes. In addition, with respect to word-counting (count of JEL codes: Duarte and Giraud 2014, Campiglio and Caruso 2007; count of terms in titles and abstracts: Panhans and Singleton 2015) it does not require the definition an a priori set of relevant terms since topic are generated by similarity in vocabulary. Finally and most important for the remainder of the study, the formation of topic is independent of the connections between citing and cited papers. This involves that mapping is not influenced by author or article popularity and that topics can encompass researchers that deal with the same

the citation frequency very close to an upward shift. So in the empirical estimation it becomes problematic to identify variations in β_2 separately from variations in α. Hence, the model contains already many parameters and (in line with Jaffe and Trajtenberg 1996 and Bacchiocchi and Montobbio 2010) we decided to concentrate our attention on the variations in α that are easier to estimate and interpret (e.g., Table 9) and we prefer not allowing variations in β_2.

6. The Blue Ribbon Eight Journals are AER, the Econometrica (ECON), the Quarterly Journal of Economics (QJE), International Economic Review (IER), JET, JPE, Review of Economic Studies (RES), Review of Economic and Statistics (RESTAT). Top journals represent the most general and advanced set of concepts that economists use in their research. So we capture the leading core of the field and the ideas and methods that are at the frontier in the leading academic institutions and have a very strong influence on the direction of research, individual careers and funding decisions. Empirically, top journals are similar in terms of impact factor, citation behavior and acceptance rates, it follows that the detection of geographical effects is less noisy than the one conducted in a more heterogeneous set of journals. Overall, top-journal knowledge is more homogeneous and general than the one contained in (top) field journals, and, therefore, more appropriate to reveal geographical patterns that only depend upon the local use of knowledge.

7. In particular, we consider the following WoS document types: "Article," "Note," "Article; Proceedings Paper"; "Proceedings Paper," "Article; Book Chapter." Other studies that use the same set of journals are Heck and Zaleski (2006) and Heck, Zaleski, and Dressler (2009), and Fourcade, Ollion, and Algan (2015).

TABLE 2
Topics Description and most Frequent Words

Topics	**Words (stemmed)**
Consumer Economics (#0)	Percent, consum, predict, day, group, advertis, sale, car, purchase, retail
Business Finance and Banks (#1)	Bank, debt, credit, borrow, patent, loan, project, entrepreneur, liquid, invest
Public Economics and Public Finance (#3)	Tax, govern, welfar, consumpt, privat, subsidi, expenditur, elast, revenu, budget
Theory of Uncertainty and Information (#4)	Agent, proof, theorem, satisfi, lemma, proposit, alloc, bid, auction, mechan
Economic Development (#5)	Region, popul, citi, land, locat, area, local, network, hous, migrat
Household Choice, Health, Insurance (#6)	Household, age, consumpt, health, insur, wealth, famili, save, care, children
Labor (#7)	Wage, worker, labor, job, unemploy, skill, earn, match, hour, search
Econometrics: Time Series (#8)	Asymptot, matrix, vector, linear, varianc, normal, regress, approxim, likelihood, econometr
Industrial Organization and Corporate Strategy (#9)	Firm, contract, profit, consum, competit, buyer, seller, incent, proposit, offer
Business Cycles and Monetary Policy (#10)	Shock, money, inflat, monetari, forecast, cycl, output, adjust, seri, nomin
International (Monetary) Economics (#11)	Countri, exchang, foreign, domest, currenc, trade, world, govern, home, bank
Portfolio Choice (#12)	Risk, asset, stock, consumpt, trade, portfolio, invest, avers, investor, uncertainty
Growth and Technology (#13)	Capit, growth, invest, output, sector, labor, industri, input, countri, elast
Game Theory (#14)	Game, player, strategi, payoff, action, belief, play, signal, learn, outcome
Education (#15)	School, educ, student, women, age, colleg, children, group, black, parent
Econometrics:Treatment Effect Models (#16)	Treatment, co, tion, match, panel, identif, heterogen, ing, outcom, bia
Corporate Governance (#17)	Firm, industri, plant, manag, coeffici, crime, regul, sale, regress, compani
Trade, Institution, Politics (#18)	Trade, tariff, export, countri, vote, voter, parti, govern, elect, candid

theme but that are not connected via co-citation and/or co-authorship.[8]

In order to extrapolate general themes, we generate 20 topics of which 18 are consistent and autonomous. The remaining two aggregate parts of the documents that do not pertain to their scientific content (such as addresses of authors or members of editorial boards): therefore, they are dropped. Finally, we consider all the documents citing our focal documents (articles, notes and proceedings papers), as reported in the Web of Science. Also in this case, we extract geographical areas from affiliations.

Following Jaffe and Trajtenberg (1996), we estimate Equation ((2) with weighted nonlinear least-squares procedure, using $\left(n_{t,a,\text{topic}}\ n_{T,A}\right)^{1/2}$ as weights. Since the left-hand variable is an empirical frequency on grouped data, the model is heteroskedastic. To improve efficiency and get the right standard errors, the weight takes into account the value of the estimated standard deviation and the observations coming from larger groups of focal and citing papers have an advantage in driving the results.

Following Jaffe and Trajtenberg (1996) and Bacchiocchi and Montobbio (2010), we also use 5-year periods for the cited years. Moreover, given that this model would return zero for lag equal to zero, we only consider cases where the citing year is strictly greater than the cited year. Finally, given that limited coverage for the citing papers at the beginning of the period, we consider only the period starting from 1990. In conclusion, we have 23 years for focal documents (1990–2012), 24 years for citing documents (1991–2015, with publication year of the citing strictly greater than the one of the focal), three areas and 18 topics. This results into a number of observations $n_{\text{obs}} = (23 \times [23 + 1]/2 + 23 \times 2) \times 3 \times 3 \times 18 = 52{,}164$.

V. PATTERNS OF GEOGRAPHICAL SPECIALIZATION AND TOPIC EVOLUTION

A. The Thematic Composition of the Top Economic Journals

Table 2 describes the topics that emerge from the sample[9] and the 10 most frequent (stemmed) words for each topic. In order to validate the LDA analysis we compare the JEL descriptors of the 10 most cited and most pertinent articles for each topic with its most frequent words to check for consistency. Results are summarized in Tables S1

8. Claveau and Gingras (2016) and Wallace, Gingras, and Duhon (2009) use such bibliometric coupling to detect themes in economics.

9. XXXX

and S2. Articles can be associated with more than one topic: the column "Weight" shows the share of topics for the listed papers.

In *Consumer Economics* (#0), Behavioral and Experimental Economics ranks among most cited and most pertinent articles in the topic. *Business Finance and Banks* (#1) partly overlaps with *Theory of Uncertainty and Information* (#4). *Public Economics and Public Finance* (#3) along with the traditional themes such as distributional effects of taxation and analysis of public policy also covers environmental issues, especially resource conservation. *Theory of Uncertainty and Information* (#4) includes general themes in microeconomics and game theory. *Economic Development* (#5) also deals with agricultural economics and economics of minorities. *Labor* (#7) focuses on wages and unemployment, while *Game Theory* (#14) mainly includes articles on bargaining theory. *Trade, Institutions and Politics* (#18) is rather heterogeneous with a stream of articles on voting behavior.[10]

This mapping is consistent both with the results Claveau and Gingras (2016, 565), especially in the relevance assigned to econometrics, and with the finding of Kosnik (2015) that signals a prevalence of microeconomic themes. Of the many novel approaches that have originated in the 1980s (Davis 2006), only experimental and behavioral economics have been able to penetrate top journals and to be an important component of a specific topic (*Consumer Economics #0*).

B. Topic Trends

Table 3 shows the evolution of topics within our focal set of documents. There is some stability of the presence of the different topics in the 28 years considered (1985–2012). However, only *Theory of Uncertainty and Information* (#4) keeps its presence constant and ranks among the most important topics at the beginning and at the end of the observed time span. In 1985, *Econometrics*: *Time Series* (#8), is the most important theme, however it undergoes a slow decline as *Treatment effects model* (#16) gains traction. In 2012, the latter is the leading topic together with #4 and the former has almost halved its relevance. A similar negative trend can be observed for the other two topics that dominate in 1985: *Industrial Organization and Corporate Strategy* (#9) and *Business Cycles and Monetary Policy* (#10). Finally, in 2012, we record an increased weight of *Economic Development* (#5), *Game Theory* (#14), and *Education* (#15).

Figure 1 shows the evolution of some topics. A substantial decrease can be noted in the importance of *Growth and Technology* (#13) and *Public Economics and Public Finance* (#3), while *Business Finance and Banks* (#1) appears to grow in importance over the whole period and especially over the last years. Finally, we show an evident switch in econometric techniques: *Econometrics and Time series* (#8) declines in 2004 leaving the lead to *Treatment effects model* (#16) that grows remarkably since 2008.

These trends confirm only partially the preexisting evidence. While corroborating the evidence on the growth of finance and economic development (Aigner et al. 2018; Kelly and Bruestle 2011), we find that the importance of industrial organization decreases as in Angrist et al. (2017).

C. Geographical Patterns and International Specialization

In what follows we use our thematic/geographical characterization of the focal set of documents to analyze the scientific profile of three geographical macro-areas: United States, Europe, and RoW. Note that in linking topics and areas we have adopted a double fractional counting because papers are assigned to more than one topic and more than one area with the appropriate weights.

Table 4 shows the number of publications in our sample for the three macro-areas. The United States cover 73% of the sample while the European share amounts to almost 16%. However, Figure 2 shows the prevalence of papers from researchers affiliated in the United States declining from 75% to less than 64% with a corresponding increase of the European share from 11% to 24% at the end of the observation period.[11]

10. The most cited articles in several topics exhibit journal clustering. The most cited papers in Consumer Economics (#0) are published prevalently in AER and in QJE; the most cited papers in Business Finance and Banks (#1) are published prevalently in the JPE and in the QJE. Theory of Uncertainty and Information (#4) and Econometrics: Time Series (#8) are concentrated in ECON while Public Economics and Public Finance (#3) is mostly present in AER and RESTAT. Finally, Economic Development (#5) clusters around AER and QJE.

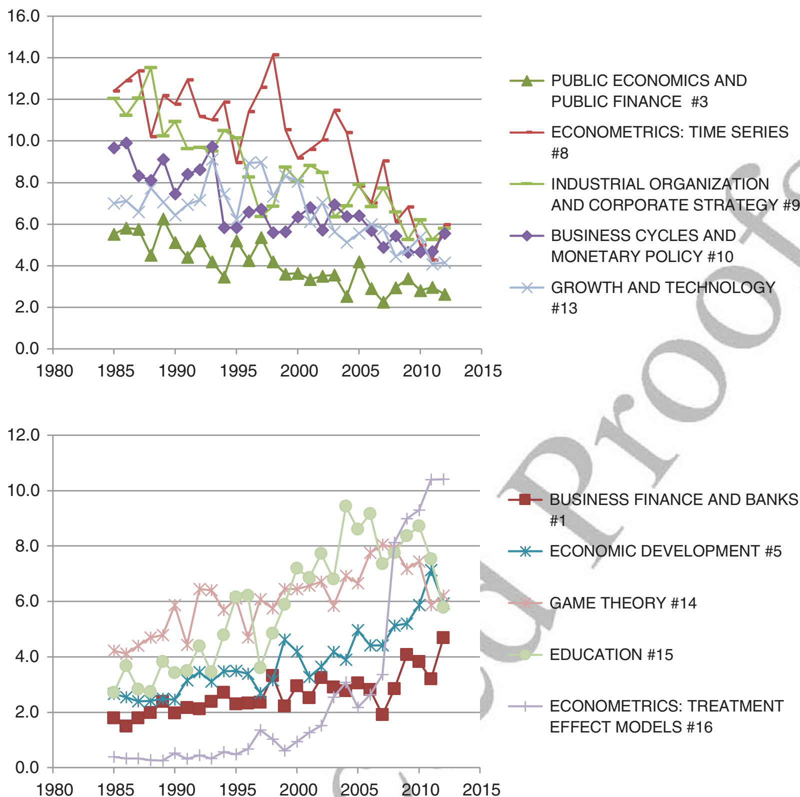
11. United Kingdom is the main contributor with approximately 30% of the European articles. However all major European countries (France, Germany, Netherlands, Spain) have experienced a growth of publications over time. For a similar trend, see Neary, Mirrlees, and Tirole (2003), Cardoso, Guimaraes, and Zimmermann (2010), Matthiessen, Schwarz, and Find (2010), and Hamermesh (2013).

TABLE 3
Distribution of Topics over Time

Topic	Consumer Economics #0	Business Finance and Banks #1	Public Economics and Public Finance #3	Theory of Uncertainty and Information #4	Economic Development #5	Household Choice, Health, Insurance #6	Labor #7	Econometrics: Time Series #8	Industrial Organization and Corporate Strategy #9	Business Cycles and Monetary Policy #10	International (Monetary) Economics #11	Portfolio Choice #12	Growth and Technology #13	Game Theory #14	Education #15	Econometrics: Treatment Effect Models #16	Corporate Governance #17	Trade, Institution, Politics #18
1985	8.4	1.8	5.5	10.9	2.7	3.2	5.6	12.4	12.0	9.7	2.8	3.4	7.0	4.2	2.7	0.4	4.7	2.6
1986	7.3	1.5	5.8	8.7	2.5	4.4	6.5	12.9	11.2	9.9	2.6	3.3	7.1	4.1	3.7	0.3	5.2	2.8
1987	7.1	1.8	5.8	11.9	2.4	4.0	6.3	13.4	12.1	8.3	2.4	3.3	6.6	4.4	2.8	0.3	4.7	2.4
1988	7.3	2.0	4.5	8.9	2.4	4.2	7.1	10.2	13.5	8.1	2.8	4.5	7.7	4.7	2.7	0.3	5.3	3.8
1989	6.1	2.4	6.3	10.0	2.5	5.4	5.1	12.2	10.3	9.1	3.1	3.3	7.1	4.8	3.8	0.3	4.6	3.5
1990	7.2	2.0	5.1	10.6	2.4	4.7	6.2	11.8	10.9	7.5	2.1	5.2	6.4	5.9	3.4	0.5	5.2	3.0
1991	8.0	2.2	4.4	10.3	3.2	4.5	6.1	12.9	9.6	8.4	3.1	3.5	6.9	4.4	3.5	0.3	5.7	3.0
1992	6.6	2.1	5.2	9.4	3.5	4.0	6.2	11.2	9.7	8.6	3.4	2.8	7.2	6.4	4.4	0.5	4.5	4.5
1993	7.1	2.4	4.2	8.8	3.1	3.4	5.3	11.0	9.5	9.7	3.5	3.7	9.1	6.4	3.4	0.3	5.5	3.4
1994	6.9	2.7	3.5	9.5	3.5	6.9	5.5	11.9	10.5	5.8	2.5	3.4	7.5	5.7	4.8	0.6	5.6	3.3
1995	8.3	2.3	5.2	9.2	3.5	6.6	5.1	9.0	10.2	5.8	2.6	3.0	6.2	6.1	6.2	0.5	5.9	4.4
1996	6.6	2.3	4.2	7.3	3.4	5.4	5.6	11.4	8.3	6.6	4.8	3.8	8.9	4.7	6.2	0.7	5.7	4.1
1997	8.2	2.3	5.4	8.3	2.7	4.6	4.6	12.6	6.4	6.7	5.4	3.2	9.0	6.1	3.6	1.4	4.8	4.9
1998	6.8	3.3	4.2	7.0	3.2	9.1	5.5	14.1	6.9	5.6	3.2	3.1	7.3	5.8	4.8	1.0	5.9	3.2
1999	6.1	2.2	3.6	10.9	4.6	5.1	7.2	10.5	8.7	5.6	3.0	2.9	8.3	6.4	5.9	0.6	4.9	3.3
2000	6.7	2.9	3.6	10.0	4.2	5.3	5.3	9.2	8.1	6.4	3.8	3.4	8.0	6.5	7.2	0.9	5.4	3.3
2001	7.9	2.5	3.4	9.6	3.3	5.2	4.6	9.6	8.8	6.8	3.5	4.3	6.1	6.6	6.8	1.3	6.3	3.3
2002	6.2	3.2	3.5	11.4	3.6	5.4	4.2	10.0	8.5	5.7	3.2	3.2	7.0	6.7	7.7	1.5	5.0	3.7
2003	7.9	2.9	3.6	7.8	4.2	7.1	4.3	11.5	6.4	7.0	3.4	4.1	5.6	5.8	6.8	2.6	5.6	3.6
2004	6.8	2.8	2.5	8.4	3.9	4.5	4.5	10.4	6.9	6.4	4.0	4.1	5.1	6.9	9.4	3.1	5.6	4.7
2005	7.8	3.1	4.2	8.6	5.0	6.2	4.7	7.8	7.9	6.4	2.8	3.4	5.6	6.7	8.6	2.2	5.0	4.2
2006	7.5	2.8	2.9	9.5	4.4	4.6	6.2	7.0	6.8	5.7	3.6	3.2	6.0	7.8	9.2	2.6	5.0	5.1
2007	7.9	1.9	2.2	9.9	4.4	6.5	4.9	9.0	7.7	4.9	2.5	4.4	5.8	8.0	7.3	3.4	4.7	4.5
2008	7.0	2.8	2.9	7.4	5.1	5.0	4.9	6.1	6.6	5.4	3.5	4.2	4.4	7.9	7.7	8.1	4.8	5.9
2009	7.8	4.1	3.4	8.1	5.2	6.8	4.6	6.8	5.3	4.6	2.6	3.7	4.8	7.2	8.4	9.0	4.4	3.4
2010	7.1	3.8	2.8	9.1	5.9	4.1	4.6	5.0	6.2	4.7	3.5	3.9	5.4	7.4	8.7	9.3	4.1	4.4
2011	10.5	3.2	3.0	8.2	7.1	5.3	4.4	4.3	5.3	4.7	2.5	3.6	4.1	5.9	7.5	10.4	5.7	4.4
2012	7.3	4.7	2.6	10.7	5.9	4.6	4.4	6.0	5.8	5.6	3.5	4.2	4.2	6.2	5.8	10.4	4.6	3.7

FIGURE 1

Evolution of Topics (Shares of Publications) in the Focal Documents: Selection of Downward Trends (Upper) and Upward Trends (Bottom)



Notably the number of publications per year attributed to the RoW is lower in the second half of the period, with a mild increasing trend over the last few years.

In order to investigate patterns of scientific specialization of these three macro areas in the 18 topics we divide the full sample (1985–2012) into two subperiods: 1985–1999 and 2000–2012, and analyze the topic profile of the scientific portfolio of different geographical areas. We build the Relative Scientific Advantage (RSA) index as the share of a topic in an area's total publication output divided by the share of this same topic over world total publication output.[12] In formal terms in each period t we calculate:

$$\mathrm{RSA}_{ik} = \frac{P_{ik} \Big/ \sum_{k=1}^{R} P_{ik}}{\sum_{i=1}^{N} P_{ik} \Big/ \sum_{i=1}^{N} \sum_{k=1}^{R} P_{ik}}$$

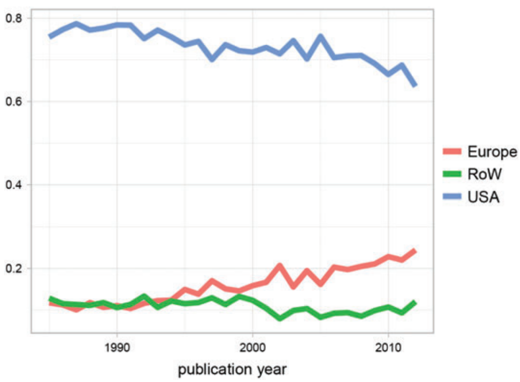
where P_{ik} is the number of publications in topic i and geographical area k. We have $R = 3$ geographical areas and $N = 18$ topics and papers are assigned to countries and topics using fractional counting. The RSA index takes values between zero and infinity. Values above one suggest a RSA (specialization). Vice versa values below one indicate a relative disadvantage (despecialization). The index is affected by a size effect because countries (in this case macro areas) with many publications are not likely to exhibit high levels of specialization. Nevertheless, some interesting facts emerge (see radar graphs in Figure 3). Overall, the U.S. publication

12. It is the traditional Balassa indicator of Revealed Comparative Advantage in international trade (Balassa 1965) applied also to innovation analysis to calculate a Revealed Technological Advantage.

TABLE 4
Publications by Geographical Areas and Overtime

Geographical Areas	Total Publications	%	1985–1999 (15 years)	Publishing/ Number of Years (a)	2000–2012 (13 years)	Publishing/ Number of Years (b)	(b) – (a)
United States	9,723.27	73.48	5,637.61	375.84	4,085.66	314.28	−61.56
Europe	2062.87	15.59	913.88	60.93	1,148.99	88.38	27.46
RoW	1,446.83	10.93	874.50	58.30	572.33	44.03	−14.27
Total	13,232.97	100.00	7,425.99	495.07	5,806.98	446.69	

FIGURE 2
Share of Publications by Geographical Areas (Papers' Affiliation)



activity is evenly distributed across topics and rather stable over time. However, it is relatively more oriented toward *Consumer Economics* (#0), *Household Choice Health insurance* (#6) *Education* (#15), *Corporate Governance* (#17), and *Business Finance and Banks* (#1).

The European and RoW areas appear substantially homogeneous in their specialization patterns. Areas of relative specialization include *Theory of Uncertainty and Information* (#4), *Econometrics – Time series* (#8), *Industrial Organization and Corporate Strategy* (#9), *Game Theory* (#14)—especially for Europe—and *Growth and Technology* (#13)—for RoW.

By comparing the two periods, Europe and RoW display a process of specialization in *Labour Economics* (#7) and *Education* (#15). The RoW also appears to have become more specialized in *Industrial Organization and Corporate Strategy* (#9). On the other hand, a process of despecialization regards *Theory of Uncertainty and Information* (#4) both in Europe and RoW.

D. *Descriptives on Citations*

Our database contains 780,180 citations. We characterize citing documents by geographical area and by publication year, which lies in the range 1985–2015.[13] Table 5 displays the geographic composition of focal and citing papers. We are comparing two very different sets of documents: seven leading economics journals, on the one side, and a less selective, much more geographically/thematically heterogeneous set on the other. As expected, compared to the focal set, the European and RoW shares of the citing papers are practically doubled.

In order to summarize how citations are distributed across topics and areas of the focal papers, we report in Table 6 an index of citation intensity. In particular, s_{to} is the share (as a percentage) of citations received by topic *to*; p_{to} is the share (as a percentage) of topic *to* in potentially cited papers; cint_{to} is the ratio $s_{\mathrm{to}}/p_{\mathrm{to}}$. *Public Economics and Public Finance* (#3), *Theory of Uncertainty and Information* (#4) and *Household Choice Health insurance* (#6) appear to be relatively less cited (citation intensity less than 1), while *Econometrics: Time Series* (#8), *Portfolio choice* (#12) and *Growth and Technology* (#13) appear relatively more cited. Similarly, s_a is the share (as a percentage) of citations received by area a[14]; p_{a} is the share (as a percentage) of area a in the set of the potentially cited papers; $\mathrm{cint}_{\mathrm{a}}$ is the ratio between s_{a} and p_{a}. It is evident that the United States attracts relatively more citations than Europe and RoW. At the same time papers originating in the RoW are relatively less cited. However, this measure might be largely influenced by the nonuniform presence over time of topics and areas in the focal documents. For

13. For our set of 780,180 forward citations we do not have the full text, so we cannot run topic modeling.

14. For example $s_{a=\mathrm{US}}$ refers to the share of citations received by papers originated in the United States. This is different from the figures in the first column of Table 5 which refer to the area of origin of the citing papers.

FIGURE 3

Specialization by Macro-Area (Balassa Index). Periods: 1985–1999 (Left) and 2000–2012 (Right)

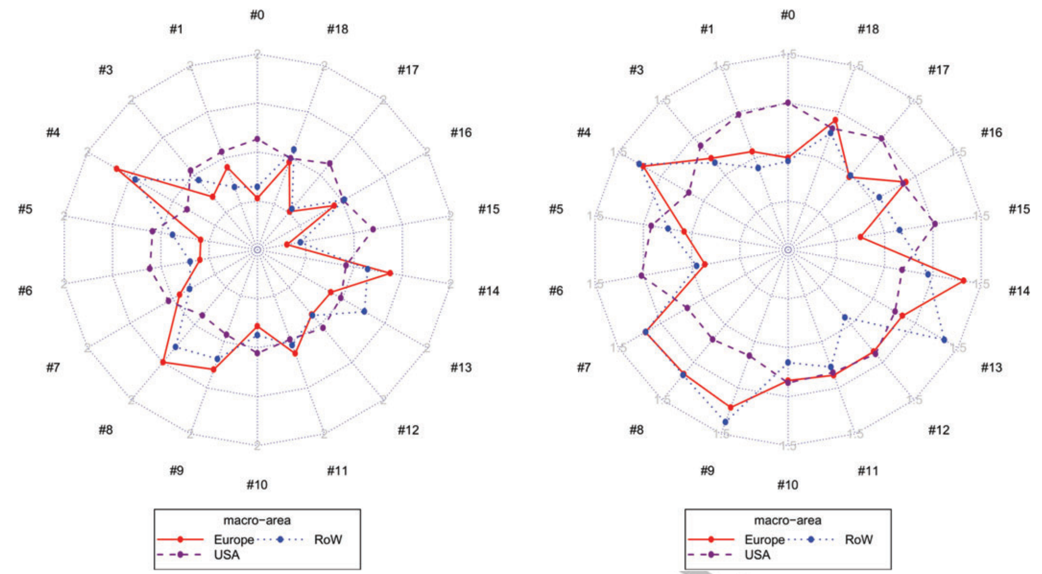


TABLE 5

Distribution by Country (in %) of Focal and Citing Documents

	Citing Papers	Focal Papers
United States	41.7	73.5
Europe	34.9	15.6
RoW	23.5	10.9

instance, since it takes time to accumulate citations, consolidated topics would have a relative advantage over recent ones in displaying high $cint_{to}$. Therefore, in order to make meaningful comparisons, we need a more structured methodology, that we present in the next section.

VI. ESTIMATION RESULTS

In this section, we report the results of the estimation of Equation ((2). The statistics for the regression variables are reported in Table 7. Table 8 displays the results. Significant tests for any particular α(k), which is a proportionality factor, focus on the null hypothesis H0: coeff =1. The null hypothesis for the significance of β_1 and β_2, instead, remains the standard H0: $\beta_i = 0$, $i = 1, 2$. A first general result regards the shapes of the citation lag distribution and the estimated values of β_1 and β_2 coefficients. The rate of decay is β_1 =0.038, while, for the rate of diffusion, the estimated value of $\beta_2 = 0.35$. As expected the rate of decay is smaller than the one observed in patent citations and the rate of diffusion is much larger (Bacchiocchi and Montobbio 2010; Jaffe and Trajtenberg 1999). These results show that the probability of being cited on average grows during the first few years, and then it decreases rather slowly as time elapses.[15] The value of the modal lag on average is about 6.7 years. The likelihood that a focal publication is cited becomes half of its estimated maximum after 28.7 years. On average after 30 years the estimated probability to be cited is still 46% of its maximum value.

To check the robustness of our results we have run the same regression on a restricted set of citing papers. In particular, we have selected the top 100 journal according to the SCImago ranking obtained from data provided by Scopus (Guerrero-Bote and Moya-Anegon 2012).[16]

15. Bjork, Offer, and Söderberg (2014) find symmetrical bell-shaped patterns of diffusion for papers written by non-Nobel winners.

16. The ranking relies on the SJR2 indicator that is computed over a journal citation network in which the nodes represent the journals, and the directed links between the nodes are the citation relationship among those journals (SCImago 2018). With respect to the IF, the SJR2 gives different weights to citations according to the prestige proximity of cited and citing journal and is size-independent.

Data are available for the period 2009–2016. Rankings do not show significant changes over time we therefore used data from 2016.

TABLE 6
Other Statistics for Thematic and Geographical Composition

Range of Focals Papers		1985–2012	
Range of citing papers		1985–2015	
Potentially cited focals		13,233	
Total citations		780,180	
Citations per potentially cited focals		59.0	
Papers by topic	s_{to}	p_{to}	$cint_{to}$
Consumer Economics (#0)	7.1	7.4	0.96
Business Finance and Banks (#1)	2.8	2.6	1.07
Public Economics and Public Finance (#3)	2.8	4.1	0.69
Theory of Uncertainty and Information (#4)	7.2	9.3	0.77
Economic Development (#5)	3.9	3.8	1.02
Household Choice, Health, Insurance (#6)	3.9	5.1	0.76
Labor (#7)	4.9	5.4	0.91
Econometrics: Time Series (#8)	13.7	10.1	1.36
Industrial Organization and Corporate Strategy (#9)	8.3	8.7	0.95
Business Cycles and Monetary Policy (#10)	6.7	6.9	0.97
International (Monetary) Economics (#11)	2.9	3.2	0.91
Portfolio Choice (#12)	4.4	3.7	1.22
Growth and Technology (#13)	8.1	6.6	1.23
Game Theory (#14)	6.4	6.0	1.08
Education (#15)	5.3	5.7	0.94
Econometrics: Treatment Effect Models (#16)	2.1	2.6	0.80
Corporate Governance (#17)	6.1	5.1	1.18
Trade, Institution, Politics (#18)	3.4	3.8	0.90
Papers by macroarea	s_a	p_a	$cint_a$
United States	78.7	73.5	1.1
Europe	13.8	15.6	0.9
RoW	7.5	10.9	0.7

Notes: $s_{to} = c_{to}/c$ and $p_{to} = n_{to}/n$, where c_{to} is number of citations by topic, n_{to} = number of (potentially cited) papers by topic, c is total number of citations, n is total number of papers, $cint_{to} = s_{to}/p_{to}$ is index of citation intensity. Similar definitions apply for s_a, p_a, and $cint_a$.

TABLE 7
Statistics for the Regression Model

Regressor	Mean	SD	Minimum	Maximum
Publication year of the focal	1997.86	5.84	1990	2012
Publication period of the focal	—	—	1990–94	2010–2013
Publication year of the citing paper	2006.93	5.98	1991	2015
Focal papers	8.55	9.32	0.06	50.65
Citing Papers	4,801.43	2,549.96	366.42	9,992.35
Citations	11.54	17.53	0.00	261.25
Lag (years)	9.07	5.98	1	25
Normalized citations (10^4)	2.98	2.54	0.00	51.18
Regression weights	167.75	111.27	5.47	711.44

Number of observations = $(23 \times [23 + 1]/2 + 23 \times 2) \times 3 \times 3 \times 18 = 52{,}164$.

Overall we have 277,000 citations with an average of 21 citations per article. Tables S3–and S5 display regression statistics and the results. It is important to note that in this case we have a much faster rate of decay: $\beta_1 = 0.073$ while the rate of diffusion is $\beta_2 = 0.38$, similar to the previous case. Accordingly, we estimate a shorter modal lag equal to 4.8 years. As a result journals with a lower ranking cite with a longer lag the journals that are higher up in the ranking.[17]

17. As an additional robustness check we have further restricted the set of citing journals to our Blue Ribbon Eight ones. In this case the estimated modal lag is 4 years. Regression results are available upon request.

TABLE 8

Estimation of Equation ((2).—α Coefficients ($N_{\text{obs}} = 52{,}164$)

Parameter	Estimate	SE	|t Value|	p Value	Significance	Parameter	Estimate	SE	|t Value|	p Value	Significance
α_{const}	1.37E-03	8.53E-05	11,711.82	< 2.2e-16	***	$\alpha_{\text{topic}=0}$	1.000	NA	NA	NA	
$\alpha_{t=1990-94}$	1.000	NA	NA	NA		$\alpha_{\text{topic}=1}$	1.045	0.027	1.63	0.103	
$\alpha_{t=1995-99}$	1.000	0.009	0.03	0.979		$\alpha_{\text{topic}=3}$	0.911	0.028	3.16	0.002	**
$\alpha_{t=2000-04}$	0.940	0.013	4.54	5.55E-06	***	$\alpha_{\text{topic}=4}$	0.914	0.024	3.53	0.000	***
$\alpha_{t=2005-09}$	0.849	0.017	8.69	< 2.2e-16	***	$\alpha_{\text{topic}=5}$	1.219	0.029	7.44	1.05E-13	***
$\alpha_{t=2010-13}$	0.726	0.022	12.62	< 2.2e-16	***	$\alpha_{\text{topic}=6}$	0.803	0.023	8.47	< 2.2e-16	***
$\alpha_{T=1991}$	1.000	NA	NA	NA		$\alpha_{\text{topic}=7}$	1.053	0.027	1.96	0.050	*
$\alpha_{T=1992}$	0.971	0.061	0.47	0.637		$\alpha_{\text{topic}=8}$	1.410	0.030	13.78	< 2.2e-16	***
$\alpha_{T=1993}$	0.851	0.051	2.89	0.004	**	$\alpha_{\text{topic}=9}$	0.908	0.023	3.95	7.94E-05	***
$\alpha_{T=1994}$	0.818	0.048	3.78	0.000	***	$\alpha_{\text{topic}=10}$	1.327	0.033	9.86	< 2.2e-16	***
$\alpha_{T=1995}$	0.752	0.044	5.65	1.63E-08	***	$\alpha_{\text{topic}=11}$	1.435	0.039	11.28	< 2.2e-16	***
$\alpha_{T=1996}$	0.719	0.042	6.71	2.01E-11	***	$\alpha_{\text{topic}=12}$	1.157	0.028	5.55	2.89E-08	***
$\alpha_{T=1997}$	0.652	0.038	9.15	< 2.2e-16	***	$\alpha_{\text{topic}=13}$	1.669	0.036	18.77	< 2.2e-16	***
$\alpha_{T=1998}$	0.610	0.036	10.93	< 2.2e-16	***	$\alpha_{\text{topic}=14}$	1.113	0.026	4.35	1.38E-05	***
$\alpha_{T=1999}$	0.581	0.034	12.27	< 2.2e-16	***	$\alpha_{\text{topic}=15}$	1.012	0.025	0.50	0.617	
$\alpha_{T=2000}$	0.578	0.034	12.43	< 2.2e-16	***	$\alpha_{\text{topic}=16}$	1.198	0.030	6.61	3.79E-11	***
$\alpha_{T=2001}$	0.548	0.032	13.96	< 2.2e-16	***	$\alpha_{\text{topic}=17}$	1.206	0.028	7.39	1.44E-13	***
$\alpha_{T=2002}$	0.537	0.032	14.53	< 2.2e-16	***	$\alpha_{\text{topic}=18}$	1.103	0.029	3.53	0.000	***
$\alpha_{T=2003}$	0.519	0.031	15.53	< 2.2e-16	***	$\alpha_{\text{focal}=\text{USA, forward}=\text{USA}}$	1.000	NA	NA	NA	
$\alpha_{T=2004}$	0.496	0.030	16.93	< 2.2e-16	***	$\alpha_{\text{focal}=\text{Eur, forward}=\text{USA}}$	0.650	0.011	31.81	< 2.2e-16	***
$\alpha_{T=2005}$	0.471	0.028	18.61	< 2.2e-16	***	$\alpha_{\text{focal}=\text{RoW, forward}=\text{USA}}$	0.595	0.013	31.77	< 2.2e-16	***
$\alpha_{T=2006}$	0.461	0.028	19.30	< 2.2e-16	***	$\alpha_{\text{focal}=\text{USA, forward}=\text{Eur}}$	0.647	0.008	45.19	< 2.2e-16	***
$\alpha_{T=2007}$	0.455	0.028	19.70	< 2.2e-16	***	$\alpha_{\text{focal}=\text{Eur, forward}=\text{Eur}}$	1.039	0.013	3.00	0.003	**
$\alpha_{T=2008}$	0.442	0.027	20.61	< 2.2e-16	***	$\alpha_{\text{focal}=\text{RoW, forward}=\text{Eur}}$	0.524	0.012	40.31	< 2.2e-16	***
$\alpha_{T=2009}$	0.424	0.026	22.00	< 2.2e-16	***	$\alpha_{\text{focal}=\text{USA, forward}=\text{RoW}}$	0.542	0.008	59.03	< 2.2e-16	***
$\alpha_{T=2010}$	0.440	0.027	20.52	< 2.2e-16	***	$\alpha_{\text{focal}=\text{Eur, forward}=\text{RoW}}$	0.525	0.011	43.82	< 2.2e-16	***
$\alpha_{T=2011}$	0.433	0.027	21.02	< 2.2e-16	***	$\alpha_{\text{focal}=\text{RoW, forward}=\text{RoW}}$	0.649	0.014	25.40	< 2.2e-16	***
$\alpha_{T=2012}$	0.419	0.026	22.01	< 2.2e-16	***						
$\alpha_{T=2013}$	0.428	0.027	21.12	< 2.2e-16	***						
$\alpha_{T=2014}$	0.421	0.027	21.52	< 2.2e-16	***						
$\alpha_{T=2015}$	0.408	0.026	22.53	< 2.2e-16	***						

Estimation of Equation ((2).—β Coefficients ($N_{\text{obs}} = 52{,}164$)

Parameter	Estimate	SE	|t Value|	p Value	Significance
$\beta_{1\ \text{const}}$	0.038	0.003	14.61	< 2.2e-16	***
$\beta_{1\ \text{topic}=0}$	1.000	NA	NA	NA	
$\beta_{1\ \text{topic}=1}$	0.739	0.073	3.57	0.000	***
$\beta_{1\ \text{topic}=3}$	1.758	0.137	5.53	3.19E-08	***
$\beta_{1\ \text{topic}=4}$	1.726	0.123	5.91	3.39E-09	***
$\beta_{1\ \text{topic}=5}$	1.080	0.081	0.99	0.324	
$\beta_{1\ \text{topic}=6}$	1.024	0.094	0.26	0.794	
$\beta_{1\ \text{topic}=7}$	1.384	0.100	3.83	0.000	***
$\beta_{1\ \text{topic}=8}$	1.525	0.095	5.50	3.80E-08	***
$\beta_{1\ \text{topic}=9}$	1.335	0.098	3.43	0.001	***
$\beta_{1\ \text{topic}=10}$	2.420	0.160	8.86	< 2.2e-16	***
$\beta_{1\ \text{topic}=11}$	2.555	0.174	8.94	< 2.2e-16	***
$\beta_{1\ \text{topic}=12}$	1.022	0.079	0.28	0.778	
$\beta_{1\ \text{topic}=13}$	1.855	0.116	7.39	1.52E-13	***
$\beta_{1\ \text{topic}=14}$	1.028	0.076	0.37	0.712	
$\beta_{1\ \text{topic}=15}$	0.728	0.068	3.99	6.55E-05	***
$\beta_{1\ \text{topic}=16}$	0.008	0.061	16.19	< 2.2e-16	***
$\beta_{1\ \text{topic}=17}$	1.020	0.075	0.27	0.788	
$\beta_{1\ \text{topic}=18}$	1.485	0.109	4.44	8.95E-06	***
$\beta_{1\ \text{focal}=\text{USA, forward}=\text{USA}}$	1.000	NA	NA	NA	
$\beta_{1\ \text{focal}=\text{Eur, forward}=\text{USA}}$	0.844	0.033	4.67	3.06E-06	***
$\beta_{1\ \text{focal}=\text{RoW, forward}=\text{USA}}$	1.206	0.049	4.20	2.66E-05	***
$\beta_{1\ \text{focal}=\text{USA, forward}=\text{Eur}}$	0.448	0.019	29.45	< 2.2e-16	***
$\beta_{1\ \text{focal}=\text{Eur, forward}=\text{Eur}}$	0.839	0.023	6.87	6.38E-12	***
$\beta_{1\ \text{focal}=\text{RoW, forward}=\text{Eur}}$	0.813	0.041	4.51	6.45E-06	***
$\beta_{1\ \text{focal}=\text{USA, forward}=\text{RoW}}$	0.327	0.021	31.62	< 2.2e-16	***
$\beta_{1\ \text{focal}=\text{Eur, forward}=\text{RoW}}$	0.422	0.032	17.79	< 2.2e-16	***
$\beta_{1\ \text{focal}=\text{RoW, forward}=\text{RoW}}$	0.921	0.041	1.94	0.052	
β_2	0.350	0.008	41.26	< 2.2e-16	***

Significant codes: 0 “***” 0.001 “**” 0.01 “*” 0.05 “.” 0.1 “ ” 1.
Significant codes: 0 “***” 0.001 “**” 0.01 “*” 0.05 “.” 0.1 “” 1.

TABLE 9

Estimated α Geographical Interaction Terms, Modal Lag, and Integral of the Curve by Cited and Citing Areas ($N_{\text{obs}} = 52{,}164$)

Complete Database				Citations Restricted to the Top 100 Journals			
	α coefficients				α coefficients		
	Citing				Citing		
Cited	**USA**	**Eur**	**RoW**	**Cited**	**USA**	**Eur**	**RoW**
USA	1.00	0.65	0.54	USA	1.00	0.68	0.64
Eur	0.65	1.04	0.52	Eur	0.67	1.10	0.58
RoW	0.59	0.52	0.65	RoW	0.54	0.52	0.83
	Modal lag				Modal lag		
	Citing				Citing		
Cited	**USA**	**Eur**	**RoW**	**Cited**	**USA**	**Eur**	**RoW**
USA	6.67	8.81	9.67	USA	4.74	5.43	5.59
Eur	7.11	7.13	8.97	Eur	4.98	4.70	5.50
RoW	6.19	7.21	6.88	RoW	4.72	4.98	4.35
	Cumulative probability (10^3)				Cumulative probability (10^3)		
	Citing				Citing		
Cited	**USA**	**Eur**	**RoW**	**Cited**	**USA**	**Eur**	**RoW**
USA	33.0	50.4	58.5	USA	18.4	17.8	18.0
Eur	25.8	41.5	43.4	Eur	13.9	19.7	15.6
RoW	16.0	21.6	23.4	RoW	9.8	10.9	12.2

A second general result refers to the estimated time effects for the citing years (α_T) and for the cited periods (α_t), that serve primarily as controls. The α_T show a downward trend that stabilizes in the last 10 years of the sample. $T = 1991$ is the base case and $\alpha_{T=1991}$ is constrained to unity, so $\alpha_{T=2004} = .49$ implies in citing year $T = 2004$ on average the probability to observe a citation is half the one observed in citing year $T = 1991$. This is because our dependent variable is the ratio $p_{t,a,\text{topic},T,A} = \frac{c_{t,a,\text{topic},T,A}}{(n_{t,a,\text{topic}})(n_{T,A})}$ and n_T grows substantially over time. So probability for the "average" citable paper to receive a citation from a paper published in $T = 2004$ (relative to $T = 1991$) is reduced due to the substantial increase in the number of potentially citing papers.

Considering the restricted citation sample, the estimated α_T are larger because there are less citing papers and, on top of this, the only difference is that the coefficients increase between 2006 ($\alpha_{T=2006} = .70$) and 2015 ($\alpha_{T=2015} = .90$). Among the restricted sample of 100 top journals we observe an increased probability to cite the Blue Ribbon Eight ones.

Finally, the coefficients for the cited period (α_t) decline steadily relative to the base (1990–1994), to .85 in 2005–2009, and .73 in 2010–2013. This downward trend suggests a decline in the observed "fertility" of publications in the most recent subperiods. A similar pattern is observed for the restricted sample where the estimated α_t are .72 in 2005–2009, and .60 in 2010–2013.

A. Geography

Table 9 reports the estimated coefficients for the interactions between geographical areas in matrix form. In particular, we report the α coefficients in the upper panel, the lag (expressed in years) at which the citation frequency reaches its maximum value in the second panel, and an estimation of the expected number of citations that a single article could potentially receive for all future years in the third panel (the precise formulas are given in Section III). The estimated α's measure the citation intensity (or "fertility" or "importance") relative to a base category. Note that for each specific category, higher values of α and higher values of β_1 (the rate of decay) would generate offsetting effects on the citation lag distribution. To understand which parameter dominates, it is therefore necessary to estimate also the overall cumulative frequencies.

Table 9 shows the estimation results for the complete database (left panel) and for the database with a restricted number of citations. In Table 9 (top panel) if we look at the data by row, the citation intensity varies with the characteristics of the citing publications and it has to be interpreted as the probability of making a citation. So we observe variation in the *use* of knowledge. As an example, if A = RoW and a = United States, then $\alpha_{aA} = .54$ means that the average publication of a scientist in the RoW is 54% as likely as a publication of a U.S. scientist to cite any given publication originated in the United States. If we look at the data by column, the citation intensity varies with the characteristics of the focal publication and it has to be interpreted as the probability of receiving a citation. So we observe variation in the importance or fertility of knowledge. So, if A = US and a = RoW, then $\alpha_{aA} = .59$ means that a publication originated in the RoW is 41% less likely to get a citation from an average U.S. publication than is a random U.S. publication.

The results (from both datasets) show clearly two overlapping forces. The first one is a home bias effect: publications whose authors are in the same geographical areas are more likely to cite each other than authors affiliated in other geographical areas. This is a pattern of geographic localization also discussed in Jaffe and Trajtenberg (1999) and Bacchiocchi and Montobbio (2010) in patent citations. The second one is a U.S. effect. Looking at the off diagonal elements, U.S. papers attract relatively more citations.

The diagonal coefficients in Table 9 (top panel) strongly dominate both the rows and columns of the matrix for the United States and Europe. In patents the localization effect seems to be stronger, it is, however, remarkable that on average a publication originated in Europe is 35% less likely to get a citation from an average U.S. publication than is a random U.S. publication. Similarly, on average a publication originated in Europe is 39% more likely to get a citation from an average European publication than is a random U.S. publication. The diagonal coefficient dominates also in the case of the RoW. However, the probability that a publication from the RoW cites another publication from the RoW is lower than the probability of a U.S.-U.S. citation. In this case the home bias effect is moderated by the heterogeneity of this group of countries.[18]

Turning the attention to the off-diagonal elements on the one side the results show the strong link between the United States and Europe, on the other side, U.S. publications seem to be more fertile: for a random European paper the probability to cite a U.S. paper is 13% (or 15%) higher than the probability to cite a paper for the RoW (65% to 52% in the left top panel and 68% to 52% in the right one). Similarly for a random RoW paper the probability to cite a U.S. paper tends to be higher than the probability to cite a paper from Europe (65% to 52%).

These results are all confirmed when we look at the results with the restricted sample, so they do not depend upon the absolute number of citations or the quality of the citing journals. The right-end side of Table 9 suggests also that the localization effect for the RoW is stronger when we consider the citations coming from the top 100 journals.

Turning the attention to the processes of diffusion and decay it is important to emphasize that in Equation ((2) both the modal lag and the cumulative probability are a negative function of the estimated β_1. With a faster decay citations come earlier and the overall number of citations is reduced. Table 9 shows that the β_1 are relatively smaller (lower obsolescence rate) when the United States is the cited country and Europe and the RoW are the citing countries. So publications originated in the United States keep on been cited in Europe and RoW for many years. On the contrary the β_1 are relatively larger (higher obsolescence rate) when Europe and the RoW are the cited countries and the United States is the citing country.

As a consequence, Table 9 shows that, in general, citations originated in the United States tend to be quicker: the first column of the second panel in Table 9 shows that, when the citing country is the United States, the values of the estimated modal lag are 6.7, 7.1 and 6.2 years for papers originated in United States, Europe, and RoW. In parallel, the modal lag is systematically higher

18. The home bias effect could be driven by the national policy relevance of the papers. So we analyzed whether the home bias effect differs between empirical and theoretical subfields. We thank a referee for pointing this out. We have exploited our topic modeling exercise to classify our topics in two groups: relatively more empirical (RME) and relatively more theoretical (RMT). We come out with a classification that is very similar to Angrist et al. (2017). We estimate the α geographical interaction terms for RMT and RME fields and we find that the home bias effect is not significantly different between them. Details are available from authors upon requests. Aside from the home-bias effect results, we find that RME fields exhibit a slower rate of decay (β_1) than RMT ones and, on average, citations to papers in RME fields have a longer modal lag. Interestingly this latter result is in line with Anauati, Galiani, and Gálvez (2016) who show that applied (and applied theory) papers have a longer life cycle of citations than theoretical papers. In our case this occurs in particular for European and RoW papers citing U.S. papers.

when the citing papers are from Europe and the RoW (see the second and the third columns). The modal lags are particularly high when there are European and RoW papers citing U.S. papers (8.8 and 9.7 years, respectively) and RoW papers citing European papers (9 years). This signals that publications in the United States get obsolete more quickly and that scientific progress advances at higher speed. These results give a precise quantitative expression to commonly held perceptions about the dynamism of the economic discipline in the United States vis-à-vis other countries. The economic discipline in the Untied States is extremely dynamic: on the one side, there are rapid developments during the first few years after an article is published and, on the other side, there is a very high rate of decay.

These results hold also for the restricted citation sample with two notable exceptions. The first one (as already noted above) is that in this case the modal lag is on average significantly shorter. The difference between the two samples is on average between 2 and 3 years. The second one is that, when only citations form high-quality journals are considered, the elements on the main diagonal are systematically lower. Citations within the same geographical area have a faster diffusion and a faster decay. However looking at the off-diagonal elements, the right-end side of Table 9 confirms that citations originating in the United States come faster.

Finally the third panel in Table 9 (bottom panel) shows the estimated cumulative probability. There are three main results. The first one is that when the cited area is the United States the values of the cumulative probabilities are systematically higher. The second one is that U.S. papers cite relatively less non-U.S. papers. The third one confirms the home bias effect in particular for Europe and the United States.

The average U.S. paper in its lifetime can expect to receive 33×10^{-3} citations[19] from a random paper (per year) originated in the United States and 58.5×10^{-3} from a paper originated in the RoW. In parallel an average paper from the RoW can expect to receive 16×10^{-3} citations from a random paper (per year) originated in the United States and 23.4×10^{-3} from a random publication in the RoW. Looking at the results by row (across columns), the estimated average number of citations is a measure of the sources of knowledge and their relative overall impact or fertility. This measure is particularly high for the United States (see the first row). This again conveys the idea of the dynamism in the United States where research has a higher impact and also where progress is very rapid. Papers from Europe and the RoW cite relatively more U.S. papers with a longer lag and this result is not affected by the database considered.

When the citing area is the United States the values of the cumulative probabilities are systematically lower. In this case looking at the Table 9 across columns, we observe variation in the use of knowledge. The estimated values in the first column are systematically lower (this holds for both the samples used in the estimations). This is because the β_I are relatively larger (higher decay rate) when the United States is the citing countries. In addition, numbers are particularly small when Europe and the RoW are the cited countries.

It is important to note that when we consider the restricted sample the result that the U.S. papers tend to cite relatively less non-U.S. papers is confirmed; however, the first row of Table 9 (right-end side of the bottom panel) shows that this does not occur the other way round. When the papers originate in the United States there are no differences in estimated cumulative probabilities across geographical areas. All the countries (considering the top 100 journals) seem to cite the U.S. papers in the same way. Conversely, when the papers originate in Europe and RoW, on the one side, they have a relatively higher probability to be cited in the same geographical area, on the other side, they receive a relatively small amount of citations from the United States. For example the average European paper in its lifetime can expect to receive 13.9×10^{-3} citations from a random paper (per year) originated in the United States, 19.7×10^{-3} citations from a random paper in Europe and, finally, 15.6×10^{-3} from a paper originated in the RoW. These results do not depend upon the overall amount of citations and it is not affected by the quality of the citing journals.

Figure 4A–C graphically shows the effects of the parameters of the matrix in Table 9. Each figure presents the estimated citation functions for citations to one of the geographical areas, with the different lines within each figure corresponding to the different citing areas. Again first of all there is evidence of geographic localization.

19. These numbers, as explained in Section III, can be considered an estimation of the expected number of citations that a single publication will receive from a set of publications consisting of one random publication per year forever. As expected these numbers are significantly larger if compared to the same estimated values for patents (Bacchiocchi and Montobbio 2010).

FIGURE 4

Fitted Curves by Citing and Cited Geographical Areas. (A) Cited Area: United States, (B) Cited Area: Europe, and (C) Cited Area: RoW

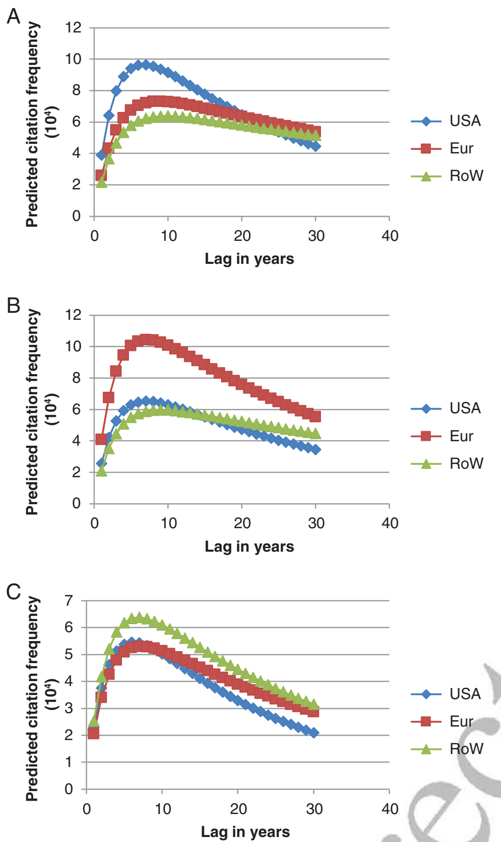


Figure 4A–4C shows that the U.S. citations to U.S. papers, European citations to European papers, and RoW citations to RoW papers are above citations across geographical areas.

Second U.S. citations come faster—as its line typically peaks early and then fades—and citations from Europe and RoW are slower. In Figure 4A the predicted frequency of citation from Europe and RoW reaches its maximum value approximately 2 and 3 years later with respect to the U.S.-U.S. case (see also Table 9, second panel first row). Figure 4A also shows that geographical localization fades away over time. The combination of relatively high α and relatively small β_1 for non-U.S. citations to U.S. publications means that the initial domestic probability is much higher, but that it fades faster, so that other countries catch up eventually.

Figure 4A shows that the U.S.-U.S. citation function crosses the other ones after 20 years. This effect is quantified in Table 10 that shows that the probability that a publication in Europe or RoW would cite—1 year after the publication date—a publication originated in the United States is 40% and 33%, respectively, lower than citations originated in the United States (42% and 39% if we consider the restricted sample), but 30 years later the figures turn out to be 21% and 16% higher (23% and 28% if we consider the restricted sample). These results measure the extent of the initial localization and the speed of fading in the United States and the lasting impact in Europe and RoW. Similarly, the relatively reduced dynamism in Europe and RoW explains why the localization effect does not fade away at the same rate for publications originated in Europe and the RoW, as shown in Figure 4B and 4C.

B. Topics

Table 8 shows the estimated values of the different α_{topic} and $\beta_{1,\,topic}$ in Equation ((2) (*topic* is an attribute of the cited papers). Thus, fields with α_{topic} larger than one are likely to get more citations than the base field (*topic* = 0) at any point in time. At the same time, the citation lag distribution of publications in topics with larger $\beta_{1,\,topic}$ have a higher degree of obsolescence. For example, $\alpha_{topic\,=\,\text{Growth and Technology}} = 1.67$ (Tables 8 and 11) means that publications in this field get on average 67% more citations as those in the base field. However, $\beta_{1,\,topic\,=\,Growth\ and\ Technology} = 1.85$ means that on average the initial amount of citations is rather large but it decays rather quickly over time. This can also be observed in Figure 5, where we plot the predicted citation function for publications in *Growth and Technology* (#13), versus publications in the other fields. Articles in *Growth and Technology* are much more highly cited during the first few years after publications; however, due to their faster obsolescence, in later years they are actually less cited than those in the base group.

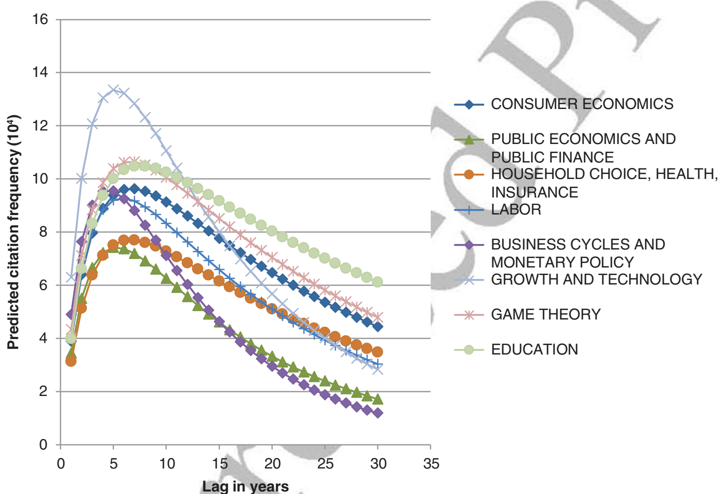
Table 11 shows the ratio of the citation probability of each topic to the citation probability of the base topic, at different lags (1, 5, 10, 20, and 30 years after the publication date of the cited article). Looking again at *Growth and Technology* (#13), the ratio starts very high at 1.62, but after 20 years it declines to 0.88, and declines further to 0.64 after 30 years. This implies that this field is extremely dynamic, with a great deal of "action"

TABLE 10
Citation Probability Ratio by Citing Geographic Area

		Complete Database				
		Lag in years				
Citing	**β_1**	**1**	**5**	**10**	**20**	**30**
USA	1.00	1.00	1.00	1.00	1.00	1.00
Eur	0.45	0.40	0.72	0.80	0.98	1.21
RoW	0.33	0.33	0.62	0.70	0.90	1.16
		Citations restricted to the top 100 journals				
		Lag in years				
Citing	**β_1**	**1**	**5**	**10**	**20**	**30**
USA	1.00	1.00	1.00	1.00	1.00	1.00
Eur	0.73	0.42	0.75	0.83	1.01	1.23
RoW	0.69	0.39	0.72	0.81	1.02	1.28

FIGURE 5
Fitted Curves by Topic

Color Figure - Online only



in the form of follow-up developments taking place during the first few years after an article is published, but also with a very high obsolescence rate. *Labor* (#7), *Econometrics: Time Series* (#8), *Business Cycles and Monetary Policy* (#10), *International (Monetary) Economics* (#11) all tend to display a similar pattern with relatively large α_{topic} and at the same time large $\beta_{1,\ topic}$.

An extreme case is the topic: *Econometrics: Treatment Effects Models* (#16). It begins at 124% of the base citation frequency, but due to the extremely low obsolescence rate after 30 years it actually stands at 366% relative to the base field. This is determined by the growing importance of this field in recent years built on a set of very influential papers of the past. *Business Finance* and *Banks and Education* display similar patterns with relatively low obsolescence rates. Note that after 30 years the ranking of fields changes substantially compared with the ranking at the beginning, suggesting that allowing for variations in both α_{topic} and $\beta_{1,\ topic}$ is important to understand the behavior of topics over time. These last three topics are also the ones with the highest predicted probabilities (Table 11), turning out to be the most influential topics after 30 years.

It is important to underline that there are some differences in the rate of obsolescence and diffusion of the different topics if we consider citations from the top 100 journals. So constraining the number of citations to a set of top journals

TABLE 11
Topic Effects: Estimated Results ($N_{\text{obs}} = 52{,}164$)

Topic	α_{topic}	$\beta_{1\ \text{topic}}$	Modal Lag	Cumulative Probability (10^3)	Citation Probability Ratio — Lag in Years: 1	5	10	20	30
Consumer Economics	1.00	1.00	6.67	32.99	1.00	1.00	1.00	1.00	1.00
Business Finance and Banks	1.04	0.74	7.46	47.86	1.05	1.10	1.15	1.27	1.40
Public Economics and Public Finance	0.91	1.76	5.26	15.92	0.89	0.79	0.69	0.52	0.39
Theory of Uncertainty and Information	0.91	1.73	5.30	16.32	0.89	0.80	0.70	0.53	0.40
Economic Development	1.22	1.08	6.47	36.93	1.22	1.20	1.18	1.15	1.11
Household Choice, Health, Insurance	0.80	1.02	6.61	25.81	0.80	0.80	0.80	0.79	0.78
Labor	1.05	1.38	5.85	24.19	1.04	0.98	0.91	0.79	0.68
Econometrics: Time Series	1.41	1.53	5.61	29.02	1.38	1.28	1.16	0.95	0.78
Industrial Organization and Corporate Strategy	0.91	1.34	5.94	21.72	0.90	0.85	0.80	0.71	0.62
Business Cycles and Monetary Policy	1.33	2.42	4.51	15.90	1.26	1.02	0.78	0.46	0.27
International (Monetary) Economics	1.43	2.56	4.39	16.10	1.35	1.07	0.80	0.45	0.25
Portfolio Choice	1.16	1.02	6.61	37.25	1.16	1.15	1.15	1.14	1.13
Growth and Technology	1.67	1.85	5.13	27.41	1.62	1.42	1.21	0.88	0.64
Game Theory	1.11	1.03	6.60	35.61	1.11	1.11	1.10	1.09	1.08
Education	1.01	0.73	7.50	47.11	1.02	1.07	1.12	1.24	1.38
Econometrics: Treatment Effect Models	1.20	0.01	20.10	5,304.76	1.24	1.44	1.74	2.52	3.66
Corporate Governance	1.21	1.02	6.62	38.91	1.20	1.20	1.20	1.19	1.18
Trade, Institution, Politics	1.10	1.49	5.67	23.40	1.08	1.01	0.92	0.77	0.64

is not neutral with respect to the pattern of diffusion by topic. Table S6 shows for example that *Corporate Governance* (#17)*, Education* (#15) and *Economic Development* (#5) display a substantial relative decrease in terms of cumulative probabilities. Other topic like *Portfolio Choice* (#12), *Business Cycles and Monetary Policy* (#10)*, Theory of Uncertainty and Information* (#4)*, International (Monetary) Economics* (#11)*, Industrial Organization and Corporate Strategy* (#9)*, Public Economics And Public Finance* (#3), and, finally*, Econometrics: Time Series* (#8)*,* display a relative increase in terms of expected lifetime citations. These results complement and extend Anauati, Galiani, and Gálvez (2016) with an important additional element: they suggest that citation-based indicators that take into account the quality of the citing journal are not neutral with respect to the topic of the papers.

C. Limitations

The results of our estimations are robust to various specifications. For our model R^2 is a poor measure of goodness of fit. In the absence of a univocal strategy for alternative measures of goodness of fit in generalized nonlinear models, we compare the empirical values of the dependent variable with the predicted ones and find that the goodness of fit is satisfactory.[20] In addition we have emphasized that topic #16 (*Econometrics: Treatment Effect Models*) is clearly behaving in a different way because there are few papers that are highly cited at the beginning of the period and the number of papers in this field grows extremely rapidly after 2005 (see Figure 1). So we have carefully checked the residuals of the model to look for the origin of the problem. We performed various diagnostic checks that indicate that the model is not fitting well those papers that display a clearly different citation history: in particular the ones that are relatively highly cited with respect to their specific *(t, a,* topic*, T, A)*-group. In fact the standard deviation of the residuals is higher for the higher quintiles of the distribution. However, it is possible to show that the problem is confined to a specific set of papers in a limited number of topics and geographical areas. In particular Figure S1, Supporting Information shows the average value of the residuals by years, topics, and geographical areas of the cited publications. The few relevant topics are displayed by column, and the different citing geographical areas by row. So problems are mainly confined to Europe and the RoW where there is a more limited number of papers and in a very specific set of years. For Europe

20. The Kolmogorov–Smirnov test suggests that, as expected, the distributions of empirical and predicted values are different but the maximum distance between the two distributions is low ($D = 0.1711$). In addition, we show that the correlation between the empirical and predicted values is high (51%) (e.g., Benšić 2015).

outliers are concentrated in Topic #16 (*Econometrics: Treatment Effects Models*) and Topic #8 *(Econometrics: Time Series)* in 1991 and for the RoW they are concentrated again (as expected) in topic #16 (*Econometrics: Treatment Effects Models*) in 1994 and 1998 and in Topic #1 (*Business Finance and Banks*) and #15 (*Education*) in 1993. To have an intuition of the phenomenon in Europe, Arellano and Bond (1991) and Johansen (1991) are two possible examples that have affected the outliers in topic #8. It is worth noting that these two papers play a role also in the outliers in topic #16 even if they enter this topic with a very small weight. Another example for the RoW in topic #16 in 1998 is Heckman et al. (1998), which enters as RoW with a weight of 0.25% because Jeffrey Andrew Smith at the time was affiliated to the University of Ontario.

D. *Discussion*

United States and European university systems have been long considered starkly different mainly as a consequence of different market conditions. Frey and Eichenberger (1993) suggested that American economists focused on more abstract topics emerging out the academic arena, whereas West European economists were used to deal with policy issues as from national and local contingencies. In parallel, the U.S. market for economists is typically considered larger, more competitive, and less regulated than the European one(s). Europe has smaller national academic markets in which different regulations and languages[21] act as barriers to competition. Such differences are hold responsible for generating the gap in dynamism and productivity between in United States and European Union (EU), and the United States advantage in the diffusion of knowledge.

In the last 40 years the EU has implemented several policies supporting research at the national and European level that have resulted in increasing output (Cardoso, Guimaraes, and Zimmermann 2010; Neary, Mirrlees, and Tirole 2003) and in a convergence toward the North American model of education and research (Borghans and Cörvers 2010). The launch of the framework programs for research (1984), the increased mobility of researchers fostered by the European Single Market (1992) and promoted by the European Research Area (2000) together with the policy elaborated according to the Lisbon strategy have made the European market for economists more homogeneous and more reactive to the worldwide increasing pressure to publish as a condition to get an academic job or a promotion (Frey et al. 2009). Currently, the process of integration is sustained by EU policy on mobility of academic staff and cross-country cooperation with the expectation that economic and cultural integration will improve productivity and quality standards (Aghion et al. 2010). The increasing share of European articles in top journals (Figure 2 and Table 4) starting from 1992 indicates a positive effect of such interventions on output delivery and corroborates the evidence on Europe catching-up with United States.

However, our analysis reveals that differences are still remarkable in the processes of knowledge diffusion and decay. Despite the increased accessibility of the products of research, guaranteed by the digitalization of scientific knowledge, our results on the geographic localization of knowledge flows (Table 9) show that national borders and, possibly, local citation networks (Thelwall and Maflahi 2015) still play a major role in directing the circulation of information (Catalini 2018).

Notwithstanding the long tradition of studies on the diffusion and networks of scientific knowledge (de Solla Price 1963), the diffusion of topics across geographical areas in economics remains quite unexplored in the literature. Our results measure the specific dynamism of the economic discipline in the United States vis-à-vis Europe and the RoW. In the United States we observe a faster rate of diffusion during the first few years after an article is published and, at the same time, a very high rate of obsolescence.

A tentative explanation of the differences between Europe and the United States involves the effect of local research traditions and of national institutional settings (Fourcade 2006) (e.g., labor market for scientists or the degree of autonomy of the university system) on the structure of communication and collaboration networks. In 2003, European economists published on average 40% of their articles in national journals[22] with a considerable heterogeneity

21. Olney (2017) also underlines that English speakers write in their native language, all the top economics journals are published in English and the quality of writing is key for success in publications. So English native speakers could have an advantage relative to nonnatives.

22. "A national journal for (a) country is a major publication outlet for authors from this country but not for authors from any other country, except possibly from a neighboring country using the same language." (Lubrano et al. 2003, 1380). In their sample none of the national journals enter the ranking of top outlets except for Economica (UK).

across countries: Austrian economists publish 6% of articles in national outlets whereas French and Italian reach 85% and 81%, respectively (Lubrano et al. 2003, Table 6, 1381).[23] These figures remark that several European countries communicate information mainly to national audiences thereby reducing the scope of knowledge circulation and the possibility to compare the scientific production of scholars across countries with a resulting friction in international mobility (Chessa et al. 2013). A similar degree of heterogeneity is found in educational programs: the share of PhD dissertations written in English (1994–2003) varies from 0% at Paris I (ETAPE) to 100% at the Universidad Autonoma of Barcelona and at the European University Institute 100% (Dréze and Estevan 2007).[24]

European and U.S. universities also exhibit substantial differences in the availability of economic resources for research activities, with a staggering advantage for United States. For instance, Harvard's annual budget corresponds to the average annual endowment assigned to the European Research Council to promote research in 25 EU countries (Dréze and Estevan 2007). The U.S. budget advantage together with a private hiring mechanism generate a degree of dynamism and competition that is not replicable in Europe where, in many cases, hiring is still regulated by national public procedures[25] and where incentives (salary and working conditions) to mobility are much lower and often nonnegotiable.

Overall, these features result in a European research network that is less connected than the U.S. one with a subsequent slowdown in the process of diffusion (Holger and Kalthaus 2018). Evidence of this phenomenon comes also from for medicine, science, and technology. The co-authorship network among the world leading research centers shows that connections are denser in United States than in Europe (Matthiessen, Schwarz, and Find 2010) with the consequence that in U.S. knowledge flows at higher speed and citations are quicker. Concerning decay, faster obsolescence can be related to faster diffusion (Caballero and Jaffe 1993) that allows a quicker exploitation and inclusion of knowledge in the production of new articles and a rapid turnover in references. As for international collaboration, Matthiessen, Schwarz, and Find (2010) emphasize that United States is less likely to make links with non-U.S. research centers, whereas collaborations within Europe are frequent. Less-frequent contacts between United States and Europe could explain why a publication originated in Europe is less likely to get a citation from an average U.S. publication independently of the publication outlet.

The connectivity of the communication network, however, is not the only determinant of knowledge diffusion. In EU, given the publication habits described above, it is likely that within country communication is dense and redundant with fast access to local knowledge and slow access to the more distant one (i.e., the network is expected to exhibit large average path length and high clustering). It has been shown that knowledge travels faster in small world networks (Beretta et al. 2018; Schilling and Phelps 2007) in which high clustering promotes local interaction and short average path-length makes distant knowledge more easily available (Chessa et al. 2013; Fleming, King, and Juda 2007; Singh 2005). In this perspective as suggested by Chessa et al. (2013), policy aiming at sustaining the mobility of researchers could not only improve quality and productivity but would also improve the speed of knowledge circulation by creating links with distant research community.

VII. CONCLUSIONS

Over the past 30 years there have been major changes in the economic discipline, in the functioning of the university system and very deep economic transformations. This paper studies the evolution of the economic discipline and the process of diffusion and decay by topic and geographical area over this long period of time (1985–2012) focusing on seven top journals that constitute the core of the field and on their forward citations. We contribute to the growing body of literature that quantitatively analyzes the evolution of the economic discipline looking at the papers' characteristics and their citation performance. We estimate precisely, using a quasi-structural model, the life cycle of the papers in economics taking into account their topic, and the

23. Belgium, Greece, Denmark, and Portugal publish about 25%–30% of articles in national journals; Spain, Germany and Ireland about 65%, Sweden and Norway about 15% and the Netherlands 8%, UK 40%.

24. Toulouse (GREMAQ) 12%, Alicante 40%, Erasmus-Rotterdam 65%, Université Catholique de Louvain 94%,

25. Although the *habilitation* is now a standard requirement for recruitment in most of the European countries, the titles needed to acquire it (quality and quantity of publications, achievements in teaching, leadership in research teams) and the institutions entitled to bestow it are not homogeneous (university, local, or national committees).

geographical origin and cohort of both citing and cited papers.

In particular, we adopt three related perspectives. The first one is the relative size and the evolution over time of the different topics. The second one is a geographic perspective and asks how the generation of scientific progress in the top journals is geographically distributed. The third one concerns the processes of diffusion and obsolescence of the newly created knowledge in economics by geographical areas and topics. We find that in the top journals in economics there is a large prevalence of articles affiliated to U.S. universities. This prevalence declines between 1985 and 2012 from 75% to 64% with a corresponding increase of the European share, which approaches one fourth of the papers at the end of the observation period. Secondly, the paper uses topic modeling to identify the evolution of topics in the discipline, quantifies the shift toward more empirical and microeconomic fields and shows the deep transformation generated by the identification revolution. In addition, topics are used to describe the scientific specialization profiles developed by the different geographical areas. Some differences emerge between geographical areas but overall we do not find a high level of international specialization and patterns of specialization are rather stable over time.

Moreover, estimating the properties of the citation lag distributions, we investigate the main features of the process of knowledge diffusion describing how citations spread over time across borders to distant locations and distinguishing the issue of speed from the issue of total intensity and impact. Our main goal is to analyze how citations to a scientific publication arrive over time, the role of the characteristics of the cited publications, and how much and how quickly different potentially citing locations absorb existing knowledge. So we estimate the shape of the citation lag distribution for different geographical areas and different topics. The modal lag on average is about 6.7 years in the entire sample and 4.8 years when we restrict the sample of the citing papers to the top 100 journals. Citations to articles in top journals in economics have a slow rate of decay. On average after 30 years the estimated probability to be cited is still 46% of its maximum value.

Our estimations quantify precisely four different and overlapping effects. Firstly, our results quantify the geographic localization of knowledge flows that we call home-bias effect. For example, a publication originated in Europe is 39% more likely to get a citation from an average European publication than is a random U.S. publication. This figure is 35% for U.S. publications. Localization effects remain important despite some evidence of an increasing importance of communication technology that greatly facilitates collaboration from a distance (Kim, Morse, and Zingales 2009). Secondly we calculate the speed at which the home-bias effect fades away over time. We find that the probability that a publication in Europe or the RoW would cite—1 year after the publication date—a publication originated in the United States is respectively 40% and 33% lower than citations originated in the United States, but 30 years later the figures turn out to be 21% and 16% higher. Third, we measure the long-lasting impact of U.S. publications on publications originated in other geographical areas. Papers from Europe and the RoW cite relatively more U.S. papers and these citations come with a longer lag. Finally, we show that in United States the field is more dynamic. On the one hand, knowledge circulates at a faster pace but, on the other, it gets rapidly old. Citations in the United States come faster and show a higher rate of decay. These results are robust to changes to the sample of the citing papers and they do not depend upon the quality of the citing journals.

Finally, we show the differences in the diffusion and impact of different topics. For example *Growth and Technology*, *Business Cycles and Monetary Policy* and *International (Monetary) Economics* are highly cited during the first years but display a quick obsolescence. High impact topics are *Econometrics: Treatment Effect Models*, *Business Finance and Banks* and *Education* which also display relatively lower obsolescence rates. *Public Economics and Public Finance* and *Theory of Uncertainty and Information* have on average a lower probability to be cited. We show that patterns of diffusion by topic display some differences changing the set of the citing journals. For example, if we constrain the number of citations to the set of 100 top journals, *Portfolio Choice* becomes a high impact topic and the impact of *Education* is reduced. This could have some important implications for citation-based indicators. In line with Anauati, Galiani, and Gálvez (2016) we show that those indicators that measure the quality of the cited journal could implicitly contain a premium for specific topics. Short-run impact factors could be larger for those topics with a faster rate of diffusion. We show also that this premium may change according to the set of citing journal considered.

This paper has a set of important limitations related to the use of the seven top journals and to the use of citations to estimate knowledge flows. We acknowledge that there is a lot of action in terms of topic development and knowledge flow outside this restricted set of journal (e.g., Anauati, Galiani, and Gálvez 2018). The use of top journals certainly implies some limitations in terms of generality of our results. An interesting next step is therefore to look at top field journals and test whether these geographical patterns are confirmed. In this direction Anauati, Galiani, and Gálvez (2018) show that citation patterns vary across journal tiers (and fields) and on average articles published in nontop five journals have a shorter life cycle. However, our paper takes a picture of the core of the discipline for those journals that affect importantly the process of recruitment and drive the evolution of the field. In addition, there are many channels of knowledge diffusion and we focus only on citations.

REFERENCES

Aghion, P., M. Dewatripont, C. Hoxby, A. Mas-Colell, and A. Sapir. "The Governance and Performance of Universities: Evidence from Europe and the US." *Economic Policy*, 25(61), 2010, 7–59.

Aigner E., F. Gloetzl, M. Aistleitner, and J. Kapeller. "The Focus of Academic Economics: Before and After the Crisis." ICAE Working Papers No. 75, Johannes Kepler University, 2018.

Anauati, V., S. Galiani, and R. H. Gálvez. "Quantifying the Life Cycle of Scholarly Articles across Fields of Economic Research." *Economic Inquiry*, 54(2), 2016, 1339–55.

———. "Differences in Citation Patterns across Journal Tiers in Economics." NBER Working Paper No. 25101, 2018.

Angrist, J., P. Azoulay, G. Ellison, R. Hill, and S. Feng Lu. "Economic Research Evolves: Fields and Styles." *American Economic Review: Papers and Proceedings*, 107(5), 2017, 293–7.

Arellano, M., and S. Bond. "Some Tests of Specification for Panel Data - Monte-Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58(2), 1991, 277–97.

Bacchiocchi, E., and F. Montobbio. "International Knowledge Diffusion and Home-Bias Effect: Do USPTO and EPO Patent Citations Tell the Same Story?" *The Scandinavian Journal of Economics*, 112(3), 2010, 441–70.

Balassa, B. "Trade Liberalisation and Revealed Comparative Advantage." *The Manchester School of Economic and Social Studies*, 33(2), 1965, 99–123.

Bardhan, P. "Journal Publication in Economics: A View from the Periphery." *Economic Journal*, 113(488), 2003, 332–7.

Benšić, M. "Properties of the Generalized Nonlinear Least Squares Method Applied for Fitting Distribution to Data." *Discussiones Mathematicae Probability and Statistics*, 35(1–2), 2015, 75–94.

Beretta, E., M. Fontana, M. Guerzoni, and A. Jordan. "Cultural Dissimilarity: Boon or Bane for Technology Diffusion?" *Technological Forecasting and Social Change*, 133, 2018, 95–103.

Bjork, S., A. Offer, and G. Söderberg. "Time Series Citation Data: The Nobel Prize in Economics." *Scientometrics*, 98(1), 2014, 185–96.

Blei, D. M., and J. D. Lafferty. "Topic models. Text mining: classification, clustering, and applications." 10(71), 2009, 34.

Blei, D. M., A. Y. Ng, and M. I. Jordan. "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, 3, 2003, 993–1022.

Borghans, L., and F. Cörvers. "Chapter 7: The Americanization of European Higher Education and Research. NBER in Clotvelter (ed.)," in *American Universities in a Global Market*. University of Chicago Press, 2010, 231–67.

Bottazzi, L., and G. Peri. "Innovation and Spillovers in Regions: Evidence from European Patent Data." *European Economic Review*, 47(4), 2003, 687–710.

Breschi, S., and F. Lissoni. "Mobility of Skilled Workers and co-Invention Networks: An Anatomy of Localized Knowledge Flows." *Journal of Economic Geography*, 9(4), 2009, 439–68.

Caballero, J. R., and A. B. Jaffe. "How High Are the Giants' Shoulders: An Empirical Assessment of Knowledge and Creative Destruction in a Model of Economic Growth," in *National Bureau of Economic Research Macroeconomics Annual*, Vol. 8, edited by O. Blanchard and S. Fisher. Cambridge, MA: MIT Press, 1993.

Campiglio, L., R. Caruso (2007). Where Economics has been headed? Multiple identity and diversity in economic literature. Evidence from top journals over the period 2000-2006. A first note. MPRA paper n. 4540.

Card, D., and S. DellaVigna. "Nine Facts about Top Journals in Economics." *Journal of Economic Literature*, 51(1), 2013, 144–61 March.

Cardoso, A., P. Guimaraes, and K. Zimmermann. "Comparing the Early Research Performance of PhD Graduates in Labor Economics in Europe and the USA." *Scientometrics*, 84(3), 2010, 621–37.

Catalini, C. "Microgeography and the Direction of Inventive Activity. Management Science." *Management Science*, 64, 2018, 3971–4470.

Cherrier, B. "Classifying Economics: A History of the JEL Codes." *Journal of Economic Literature*, 55(2), 2017, 545–79.

Chessa, A., A. Morescalchi, F. Pammolli, O. Penner, A. M. Petersen, and M. Riccaboni. "Is Europe Evolving toward an Integrated Research Area?" *Science*, 339(6120), 2013, 650–1.

Claveau, F., and Y. Gingras. "Macrodynamics of Economics: A Bibliometric History." *History of Political Economy*, 48(4), 2016, 551–92.

Conroy, M. E., and R. Dusansky. "The Productivity of Economics Departments in the US: Publications in Core Journals." *Journal of Economic Literature*, 33, 1995, 1966–71.

Criscuolo, P., and B. Verspagen. "Does it Matter where Patent Citations Come from? Inventor Vs. Examiner Citations in European Patents." *Research Policy*, 37(10), 2008, 1892–908.

Das, J., Q. Toan Do, K. Shaines, and S. Srinivasan. "U.S. and Them: The Geography of Academic Research." *Journal of Development Economics*, 105, 2013, 112–30.

Dasgupta, P., and P. A. David. "Towards a New Economics of Science." *Research Policy*, 23(5), 1994, 487–521.

Davis, J. B. "The Turn in Economics: Neoclassical Dominance to Mainstream Pluralism?" *Journal of Institutional Economics*, 2(01), 2006, 1–20.

Dréze, J. H., and F. Estevan. "Research and Higher Education in Economics: Can We Deliver the Lisbon Objectives?" *Journal of the European Economic Association*, 5(2-3), 2007, 271–304.

Duarte, P., and Y. Giraud. "Chasing the B: A Bibliographic Account of Economics Relation to its Past, 1991-2011." THEMA Working Paper No 2014-09, 2014

Ellison, G. "How Does the Market Use Citation Data? The Hirsch Index in Economics." *American Economic Journal: Applied Economics*, 5(3), 2013, 63–90.

Fleming, L., C. King, and A. I. Juda. "Small Worlds and Regional Innovation." *Organization Science*, 18(6), 2007, 938–54.

Fourcade, M. "The Construction of a Global Profession: The Transnationalization of Economics." *American Journal of Sociology*, 112(1), 2006, 145–94. https://doi.org/10.1086/502693.

Fourcade, M., E. Ollion, and Y. Algan. "The Superiority of Economists." *Journal of Economic Perspectives*, 29(1), 2015, 89–114.

Frey, B. S., and R. Eichenberger. "American and European Economics and Economists." *Journal of Economic Perspectives*, 7(4), 1993, 185–93.

Galiani S., and R. H. Gálvez "The Life Cycle of Scholarly Articles across Fields of Research." NBER Working Paper No. 23447, 2017

Gibson, J., D. L. Anderson, and J. Tressler. "Which Journal Rankings Best Explain Academic Salaries? Evidence from the University of California." *Economic Inquiry*, 52(4), 2014, 1322–40.

Gittelman, M., and B. Kogut. "Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns." *Management Science*, 49(4), 2003, 366–82.

Grossman, G., and E. Helpman. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press, 1991.

Guerrero-Bote, V., and F. Moya-Anegon. "A Further Step Forward in Measuring Journals' Scientific Prestige: TheSJR2 Indicator." *Journal of Informetrics*, 6(4), 2012, 674–88.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper no. 8498, 2001.

———. "Market Value and Patent Citations." *RAND Journal of Economics*, 36, 2005, 16–38.

Hamermesh, D. S. "Six Decades of Top Economics Publishing: Who and How?" *Journal of Economic Literature*, 51(1), 2013, 162–72.

——— "Citations in Economics: Measurement, Uses, and Impacts." *Journal of Economic Literature*, 56(1), 2018, 115–56.

Hamermesh, D. S., and G. A. Pfann. "Reputation and Earnings: The Roles of Quality and Quantity in Academe." *Economic Inquiry*, 50(1), 2012, 1–16.

Hargreaves Heap, S. P., and A. Parikh. "The Diffusion of Ideas in the Academy: A Quantitative Illustration from Economics." *Research Policy*, 34(10), 2005, 1619–32.

Heck, J. L., and P. A. Zaleski. "The most Frequent Contributors to the Elite Economics Journals: Half Century of Contributions to the "Blue Ribbon Eight"." *Journal of Economics and Finance*, 30(1), 2006, 1–37.

Heck, J. L., P. A. Zaleski, and S. J. Dressler. "Leading Institutional Contributors to the Elite Economic Journals." *Applied Economics*, 41(17), 2009, 2191–219.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. "Characterizing Selection Bias Using Experimental Data." *Econometrica*, 66(5), 1998, 1017–98.

Hoffman, M., Bach, F.R., Blei, D M. "Online Learning for Latent Dirichlet Allocation. Paper presented at the Advances in Neural Information Processing Systems, Vancouver, Canada, 2010, 856–864.

Holger, G., and M. Kalthaus. "International Research Networks: Determinants of Country Embeddedness." *Research Policy*, 47(7), 2018, 1198–214.

Jaffe, A. B., and M. Trajtenberg. "Flow of Knowledge from Universities and Federal Laboratories: Modelling the Flow of Patent Citations over Time and across Institutional and Geographic Boundaries." *Proceedings of the National Academy of Sciences*, 93(23), 1996, 12671–7.

———. "International Knowledge Flows: Evidence from Patent Citations." *Economics of Innovation and New Technology*, 8(1-2), 1999, 105–36.

Jaffe, A. B., M. Trajtenberg, and R. Henderson. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics*, 108(3), 1993, 577–98.

Johansen, S. "Estimation and Hypothesis-Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica*, 59(6), 1991, 1551–80.

Kalaitzidakis, P., T. Mamuneas, A. Savvides, and T. Stengos. "Research Spillovers among European and North-American Economics Departments." *Economics of Education Review*, 23(2), 2004, 191–202.

Kelly, M. A., and S. Bruestle. "Trend of Subjects Published in Economics Journals 1969–2007." *Economic Inquiry*, 49(3), 2011, 658–73.

Kim, E. H., A. Morse, and L. Zingales. "What Has Mattered to Economics since 1970." *Journal of Economic Perspectives*, 20(4), 2006, 189–202.

———. "Are Elite Universities Losing their Competitive Edge?" *Journal of Financial Economics*, 93(3), 2009, 353–81.

Kosnik, L. R. "What Have Economists Been Doing for the Last 50 Years? A Text Analysis of Published Academic Research from 1960–2010." *Economics: The Open-Access, Open-Assessment E-Journal*, 9(2015–13), 2015, 1–38. https://doi.org/10.5018/economics-ejournal.ja.2015-13.

Kuhn, T. S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

Lubrano, M., L. Bauwens, A. Kirman, and C. Protopopescu. "Ranking Economics Departments in Europe: A Statistical Approach." *Journal of the European Economic Association*, 1(6), 2003, 1367–401.

Maruseth, P. B., and B. Verspagen. "Knowledge Spillovers in Europe: A Patent Citations Analysis." *Scandinavian Journal of Economics*, 104(4), 2002, 531–45.

Matthiessen, C. W., A. W. Schwarz, and S. Find. "World Cities of Scientific Knowledge: Systems, Networks and Potential Dynamics: An Analysis Based on Bibliometric Indicators." *Urban Studies*, 47(9), 2010, 1879–97.

Merton, R. K. "The Matthew Effect in Science." *Science*, 159(3810), 1968, 56–63.

Narin, F., K. S. Hamilton, and D. Olivastro. "The Increasing Linkage between U.S. Technology and Public Science." *Research Policy*, 26(3), 1997, 17–330.

Neary, J. P., J. A. Mirrlees, and J. Tirole. "Evaluating Economics Research in Europe: An Introduction." *Journal of the European Economic Association*, 1(6), 2003, 1239–49.

Olney, W. W. "English Proficiency and Labor Market Performance: Evidence from the Economics Profession." *Economic Inquiry*, 55(1), 2017, 202–22.

Panhans, M.T, and J.D. Singleton. "The Empirical Economist's Toolkit: From Models to Methods." Center for the History of Political Economy (CHOPE) Working Paper No. 2015–03, 2015

Peri, G. "Determinants of Knowledge Flows and their Effect on Innovation." *Review of Economics and Statistics*, 87(2), 2005, 308–22.

Phelps, E. S. "Models of Technical Progress and the Golden Rule of Research." *Review of Economic Studies*, 33(2), 1996, 133–45.

Řehůřek, R., and P. Sojka. *Software Framework for Topic Modelling with Large Corpora*. Malta, 45–50: University of Malta, 2010. https://doi.org/10.13140/2.1.2393.1847.

Romer, P. M. "Endogenous Technological Change." *Journal of Political Economy*, 98(5), 1991, S71–S102.

Schilling, M., and C. C. Phelps. "Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation." *Management Science*, 53(7), 2007, 1113–26.

SCImago, "SJR — SCImago Journal & Country Rank [Portal]." 2018. Accessed March 2018, http://www.scimagojr.com.

Singh, J. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science*, 51(5), 2005, 756–70.

de Solla Price, D. J. *Little Science Big Science*. Columbia University Press, 1963.

Stephan, P. *How Economics Shapes Science*. Harvard University Press, 2012.

Thelwall, M., and N. Maflahi. "Are Scholarly Articles Disproportionately Read in their Own Country? An Analysis of Mendeley Readers." *Journal of the Association for Information Science and Technology*, 66(6), 2015, 1124–35.

Trajtenberg, M. "A Penny for your Quotes: Patent Citations and the Value of Innovations." *RAND Journal of Economics*, 21(1), 1990, 172–87.

Wallace, M. L., Y. Gingras, and R. Duhon. "A New Approach for Detecting Scientific Specialties from Raw Co-Citation Networks." *Journal of the American Society for Information Science and Technology*, 60(2), 2009, 240–6.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Plot of the residuals by year, topic and country of the cited papers

Table S1. Topics, with most pertinent documents and their JEL codes

Table S2. Topics, with most cited documents and their JEL codes.

Table S3. Statistics for the regression model – Top 100 journals

Table S4. Estimation of Equation (2 – α coefficients, Top 100 journals

Table S5. Estimation of Equation (2 – β coefficients, Top 100 journals

Table S6. Topics effects: estimated results – Top 100 journals